

SCHAUM'S
OUTLINE
SERIES

THEORY and PROBLEMS

ABSTRACT ALGEBRA

by JOONG FANG

including
785
solved
problems

Completely Solved in Detail

•
SCHAUM PUBLISHING CO.

NEW YORK

AP
FAN

10V ERSIZE

281-
P7

SCHAUM'S OUTLINE OF
THEORY AND PROBLEMS

of

ABSTRACT
ALGEBRA

'35

1965



BY

JOONG FANG, Ph.D.

Department of Mathematics

Northern Illinois University

SCHAUM PUBLISHING CO.

Sole Agents for the United Kingdom

H. JONAS & CO. (BOOKS) LTD.

18, Bruton Place, Berkeley Square,
London, W.1.

COPYRIGHT © 1963, BY THE
SCHAUM PUBLISHING COMPANY

*All rights reserved. This book or any
part thereof may not be reproduced in
any form without written permission
from the publishers.*

PRINTED IN THE UNITED STATES OF AMERICA

Typography by Signs and Symbols, Inc., New York, N. Y.

Preface

This book is designed for use either as a supplement to all current standard textbooks or as a textbook for a formal course in abstract algebra. It aims, above all, at an organic unity of the axiomatic structure of elementary abstract algebra at the sophomore, junior and possibly senior level, which will lead toward more advanced studies in this and related fields. It treats, therefore, only "basic concepts" of abstract algebra such that some, but certainly not all, fundamental results in classic and modern algebras will find their due place here. Matrices, for instance, makes only a brief appearance here as a fundamental concept, viz. as an example of noncommutative rings, and its further development is left to an independent work, *Linear Algebra*, which will be published as a sequence to the present volume.

Some early authors in this field attempted, perhaps not always successfully, to illustrate new abstract concepts in terms of as many familiar examples as possible from the classic theory of numbers and equations. Given a limited space, however, they could not but be circumspect in the choice of the most fitting topics. For, after all, abstract algebra is no substitute for the theory of numbers and equations in entirety, a full treatment of which should be carried out separately. However, some substantial parts of these topics do appear in this text.

A renewed emphasis should be put on the self-evident, but often neglected, dictum that the abstract is vacuous without the concrete. "But abstract theorems are empty words", wrote Professor C. C. MacDuffee two decades ago, "to those who are not familiar with the concrete facts which they generalize. One of the major problems in teaching abstract algebra is to give to the student a selected body of facts from number theory, group theory, etc., so that he will have the background to understand and appreciate the generalized results. Without this background, the game of playing with postulates becomes absurd." This is even more true today, especially at the sophomore and junior levels. The beginner should be properly warned against "biting off more than he can chew".

In this spirit the present book does try to bring in as many small but "chewy" topics as possible within the scope of its self-imposed limitation. As such, it is divided into five parts: *Algebra of Logic*, *Algebra of Sets*, *Algebra of Groups*, *Algebra of Rings*, *Algebra of Fields*. Each part may be studied independently, although the parts are all interdependent as an organic whole; this latter feature is manifest in an almost excessive use of cross-reference throughout the work.

Logical sequence is the guiding principle in every part of this book. Integers, for instance, get proper attention at a later stage, contrary to the traditional works, because they are considered here within the frame of integral domains, which in turn appear only after the introduction of commutative rings. Since the improving freshman courses in the last decade have absorbed much material once taught at the start of abstract algebra, a certain amount of knowledge on the domain of integers and the familiar number fields in terms of algebraic systems is taken for granted from the very beginning. This book certainly does not pretend to build up the whole structure of modern algebra from the most primitive concepts — a task comparable to that of creating something out of nothing.

Not, however, that this book is not "self-contained". As a matter of fact, every theorem within its reach is introduced here, at times with secondary proofs, except for a few rather difficult theorems which need elaborate lemmata and unproportionately many pages, such as an essentially algebraic proof of the so-called fundamental theorem of algebra and Abel's proof on the algebraic insolubility of quintic equations. The student who uses this book will seldom be in need of consulting other sources for basic theorems.

Every problem, except supplementary problems, is proved or solved on the strength of the theorems which are proved here. The student who consults this book only to find proofs or solutions for his specific problems is warned at the start that he should be quite clearly aware of the pitfalls he may encounter. For, first of all, symbols may represent different algebraic concepts, and the *context* in which the proofs or solutions are carried out here may be different from that of the textbook he uses in class. In such cases some modifications will be called for, which will be left to the student. The task of modifications, or acclimatization in general, should be well within the student's scope, since he is assumed here, as a sophomore at least, to have mastered College Algebra and some earlier parts of elementary Calculus with Analytic Geometry. The Table of Symbols, which follows the Introduction, will be of some help to the student, particularly in the period of initiation.

Thanks are due my teachers and friends for their generous interest in my work: Mr. H. Simpson, formerly Dean of Yale University Graduate School; Professor W. Kalinowski of St. John's University; Professors T. Chorbajian, J. O. Distad, F. D. Parker, D. R. Simpson, and D. Coonfield of University of Alaska; and Professor E. W. Hellmich of Northern Illinois University. Particular thanks are extended to the staff of the Schaum Publishing Company for their valuable suggestions and most helpful cooperation.

J. FANG

Northern Illinois University
March, 1963

CONTENTS

Introduction

Table of Symbols

Part 1 – Algebra of Logic

Chapter 1.1 MATHEMATICAL LOGIC.....	1
1.1.1 Tautologies	1
*1.1.2 Quantifications	13
Chapter *1.2 MATHEMATICAL PROOFS.....	19
Supplementary Problems	22

Part 2 – Algebra of Sets

Chapter 2.1 SETS IN GENERAL.....	24
Chapter 2.2 OPERATIONS.....	31
2.2.1 Operations in General	31
2.2.2 Transformations	34
Chapter 2.3 OPERATIONS ON SETS.....	40
Chapter 2.4 ABSTRACT STRUCTURES	49
*2.4.1 Lattices	49
2.4.2 Boolean Algebras	56
Supplementary Problems	63

Part 3 – Algebra of Groups

Chapter 3.1 FINITE GROUPS	65
3.1.1 Groups in General	65
3.1.2 Groups of Permutations	72
3.1.3 Homomorphism and Isomorphism	83

CONTENTS

Chapter 3.2 SUBGROUPS	90
3.2.1 Cyclic Subgroups	90
3.2.2 Cosets and Conjugates	95
*3.2.3 Normalizers and Centralizers	101
*3.2.4 Endomorphism and Automorphism....	105
*3.2.5 Normal Subgroups	110
*3.2.6 Quotient Groups	115
*3.2.7 Composition Series and Direct Products	122
Supplementary Problems	128

Part 4 – Algebra of Rings

Chapter 4.1 RINGS	131
4.1.1 Rings in General	131
4.1.2 Commutative Rings	139
4.1.2.1 Boolean Rings	139
4.1.2.2 Integral Domains	141
4.1.2.3 Integers	146
4.1.2.4 Fields in General	159
4.1.2.5 Polynomials in General	165
4.1.3 Noncommutative Rings	175
4.1.3.1 Sfields and Quaternions	175
4.1.3.2 Matrices	179
Chapter *4.2 SUBRINGS	198
*4.2.1 Subrings in General	198
*4.2.2 Ideals	201
*4.2.3 Quotient Rings	205
Supplementary Problems	210

Part 5 – Algebra of Fields

Chapter 5.1 NUMBER FIELDS	214
5.1.1 Rational Numbers	214
5.1.2 Real Numbers	219
5.1.3 Complex Numbers	236
Chapter 5.2 POLYNOMIALS OVER FIELDS	251
5.2.1 Irreducible Polynomials	251
5.2.2 Symmetric Polynomials	270
5.2.3 Roots of Polynomials	280
Chapter *5.3 ALGEBRAIC FIELDS	301
*5.3.1 Algebraic Extensions	301
*5.3.2 Algebraic Numbers	311
Supplementary Problems	321
ANSWERS AND HINTS	325
INDEX	335

Introduction

The student is advised to make use of the cross-references in every part of the book, and of the Table of Symbols following this Introduction and of the Index at the end of the book. The cross-references are usually given in the form "cf. Th.2.2.2.16", for example, meaning "refer to the theorem, numbered 16, in Part 2, Chapter 2, Section 2". "Df.", "Prob.", and "MTh." denote a "definition", a "solved problem", and a "metatheorem" (i.e. theorem of theorems, which is not to be proved in terms of ordinary definitions and theorems) respectively. Such cross-references should be consulted as often and carefully as possible, since they indicate the reasoning or justification behind the steps of proofs or solutions.

Starred definitions, theorems and problems are optional; they may be skipped in the first reading, although they may still be referred to in the subsequent sections. All metatheorems are starred in principle, since they cannot be proved properly within the frame of the main text, although they are quite freely adapted here.

Boldface letters and Greek letters are used very sparingly, indeed only when absolutely necessary. Script letters and Hebrew letters are not employed in the text for an elementary reason: there are too few letters, in whichever form or language, to permit every algebraic concept or system monopolize a certain type of letters. There are, and will be, too many novel ideas in mathematics to be exhaustively and mutually exclusively classified by a few types of letters.

The student, then, must learn as early as possible to decipher the meaning of what few letters he has *within a certain context*. The context, and not merely the type of letters, is to yield a coherent and consistent meaning of the text. " R ", for instance, may designate "a ring" here and "the rational number field" there, but it will not at all confuse the student if he thinks of the context before everything else.

In the same spirit such terms as "module" or "complex" are used quite freely, taking the risk of incurring the purist's wrath. The liberalism with respect to symbols and terms may be considered a part of mathematical training, however, since the student must face similar situations sooner or later. The student at the sophomore or junior level may be, or rather should be, expected to be able to distinguish the " I " representing "an identity mapping" from the " I " denoting "the domain of integers" in two different contexts. Such a training may be considered quite pertinent or even essential, in abstract algebra in particular. For, after all, abstract algebra was born through the awareness of a unifying theory under the existence of parallel theories in many branches of classic algebra. The student should be encouraged to learn such characteristics in mathematical reasoning as soon as he is ready to pursue the fascinating enterprise.

Reasoning in general may transcend a certain logic, but mathematical reasoning cannot; it is, in its written form at least, confined within the frame of mathematical logic. Hence the study begins with Algebra of Logic. Because of the severely limited scope of the book, however, it barely scratches the surface of the profound subject, allowing the student only a bird's-eye view. The interested student may pursue the subject in the following readily available book:

Langer, S. K., *An Introduction to Symbolic Logic*, 2nd Ed., Dover, 1953

Algebra of Logic is followed by Part 2, Algebra of Sets, without which no modern mathematics can begin. Again, because of the limited scope and space, only an elementary theory of sets is presented, leaving a supplementary and more advanced study to the following books:

- Birkhoff, G., *Lattice Theory*, 2nd Ed., A.M.S. Colloquium, vol. 25, 1948
- Chevalley, C., *Fundamental Concepts of Algebra*, Academic, 1957
- Dieudonné, J., *Foundations of Modern Analysis*, esp. Chap. 1, Academic, 1960
- Hamilton, N. T., and Landon, J., *Set Theory*, Allyn and Bacon, 1961
- Hohn, F., *Applied Boolean Algebra*, Macmillan, 1960
- Kamke, E., *Theory of Sets*, Dover, 1950

It must be noted that the new terms “injective”, “surjective”, and “bijective” with respect to mappings in §2.2.2 closely follow Dieudonné’s work.

Part 3, Algebra of Groups, is an elementary presentation of the theory of finite groups. This is a well-explored field, which as such is abundant in literature. The following list, then, is merely a representative one for the beginner:

- Alexandroff, P. S., *An Introduction to the Theory of Groups*, Hafner, 1959
- Hall, M., *The Theory of Groups*, Macmillan, 1959
- Kurosh, A., *Theory of Groups*, 2 vols., Chelsea
- Ledermann, W., *The Theory of Finite Groups*, Interscience, 1953
- Zassenhaus, H., *The Theory of Groups*, 2nd Ed., Chelsea, 1956

Part 4, Algebra of Rings, and Part 5, Algebra of Fields, are so closely related at this elementary level that they may share the following bibliography in common:

- Albert, A. A., *Fundamental Concepts of Higher Algebra*, U. of Chicago, 1956
- Borofsky, S., *Elementary Theory of Equations*, Macmillan, 1950
- Jacobson, N., *Structure of Rings*, A.M.S., 1956
- McCoy, N. H., *Rings and Ideals*, M.A.A., 1948
- Pollard, H., *The Theory of Algebraic Numbers*, M.A.A., 1950
- Uspensky, J. V., *Theory of Equations*, McGraw-Hill, 1948
- Van der Waerden, B., *Modern Algebra*, 2 vols., Unger, 1949-50
- Weisner, L., *Introduction to the Theory of Equations*, Macmillan, 1938
- Weyl, H., *Algebraic Theory of Numbers*, Princeton, 1940

At the end of each part there appears a collection of supplementary problems, most of which are to sharpen the student’s skill in solving problems, possibly providing additional detail about the material covered in the main text. The student who wishes to master the subject should solve a good many of these by his own efforts, although he should not be disheartened if he cannot solve all of them by himself. Some of these, the starred ones in particular, are rather difficult, and the student should better leave them alone, for the time being at least, until he masters the ways of reasoning in the solved problems. For the ambitious, however, “the sky is the limit,” and the student is invited to be as ambitious as possible.

Table of Symbols

Df.	Definition.
Th.	Theorem.
MTh.	Metatheorem.
Prob.	Problem (solved).
Hyp.	Hypothesis.
\therefore	Hence.
\because	Since.
i.e.	That is.
viz.	Namely.
iff	If and only if.
$\vdash p$	Yields p (assertion).
\bar{p}	Not p .
$p \cdot q$ (or $pq, p \wedge q$)	p and q .
$p \vee q$	p or q .
$p \leq q$	p or q but not both.
$p \mid q$	Not p or not q (or: not both p and q).
$p \downarrow q$	Neither p nor q .
$p \rightarrow q$	If p , then q (or, p implies q ; or, p only if q).
$p \leftrightarrow q$ (or $p \equiv q$)	p iff q .
$(\exists x)(\dots)$	There exists x such that ...
$(x)(\dots)$ or $\{x \mid \dots\}$ or $\{x: \dots\}$	For all x such that ...
\circ (or $*$)	An operator in a postulational algebraic system, with $x \circ y$ (or $x * y$) as an element of the system.
A, B, C , etc.	The boldface italic capital letters denote classes, i.e. collections of sets, which should be distinguished from the sets in themselves.
G1, G2 , etc.	The boldface Roman capital letters with numbers are to number the postulates for a certain algebraic system; G1 , then, denotes the first postulate to characterize the concept of groups, and G2' , for instance, designates the second postulate of the second alternative set of postulates for groups. Likewise, G4'' denotes the fourth postulate of the third alternative set of axioms for groups. Further examples are:
P1, P2, ..., P5	Five tautologies of the Principia Mathematica.
L1, L2, ..., L4	Four axioms which characterize a lattice.
O1, O2, ..., O4	Four axioms of ordering.
B1, B2, ..., B6	Six postulates for a Boolean algebra.
R1, R2, ..., R8	Eight postulates for a ring.
$\bar{B}1, \bar{B}2, \dots, \bar{B}9$	Nine postulates for a Boolean ring.
D1, D2, ..., D11	Eleven postulates for an integral domain.
N1, N2, ..., N4	Four axioms for the set N of natural numbers.
F1, F2, ..., F11	Eleven postulates for a field.
V1, V2, ..., V8	Eight postulates for a vector space.

TABLE OF SYMBOLS

A, B, C, \dots, X, Y, Z	Light faced italic capital letters denote sets in <i>most</i> cases; otherwise, specifications will be explicitly given in the context. Of these capital letters, some will <i>almost always</i> designate certain sets in particular, although they are by no means monopolized by some specific sets on all occasions (cf. Introduction). Typical cases are:
A	A total matric algebra.
B	A Boolean algebra.
C	The complex number field.
D	An integral domain.
D^+	A complex (non-empty subset) of D containing only positive, or more generally, non-negative elements.
\bar{D}	An ordered integral domain.
F^+	A complex of F containing only non-negative elements.
\bar{F}	An ordered field.
F^*	A sfield (or division ring).
G	A group.
\bar{G}	The field of all Gaussian numbers.
\bar{G}'	The algebraic number field of all Gaussian integers.
I (or J)	The integral domain of integers (or rational integers).
I^+	A complex of I containing only non-negative elements.
I_+	I regarded as a group under addition.
\bar{I}	The integral domain of all algebraic integers.
I_m (or $I/\{m\}$, or $I/(m)$)	The residue classes of integers modulo m .
J	The same as I , replacing I now and then, when I denotes an identity mapping, and in particular when the feature of Df. 4.1.2.3.5 with respect to I is stressed.
L	A lattice.
N	The set of natural numbers.
\emptyset	The null (or vacuous or empty) set.
P_n	A permutation group of order n .
Q	A quotient field.
\bar{Q}	The sfield of quaternions.
R	The rational number field (or a ring in general).
R^+	A complex of R containing only non-negative elements.
\bar{R} (or R^*)	The real number field.
\bar{R}^+	A complex of \bar{R} containing only non-negative elements.
S_n	A symmetric group of order n .
V (or $V(R)$, etc.)	A vector space (over R , etc.).
V_4	Klein's (or "four") group, i.e. the so-called "Vierergruppe".
$\alpha, \beta, \gamma, \dots$	Vectors.
a, b, c, \dots, x, y, z	Small letters generally denote the elements of a set.
$\{a, b, \dots\}$	a, b, \dots as listed elements.
$a \in A$	The element a belonging to the set A .
$a \notin A$	The element a not belonging to the set A .
A'	The complement of A .
$A - B$	The complement of B in A .
$A \times B$	The Cartesian (or direct) product of A and B .

TABLE OF SYMBOLS

$X \subset Y$	X is a (proper) subset of Y .
$X \subseteq Y$	X is a subset of Y .
$X \cup Y$	The join (or logical sum or union) of X and Y .
$X \cap Y$	The meet (or logical product or intersection) of X and Y .
$\cup_{a \in A} S_a$	The set of all elements which belong to S for some a of A .
$\cap_{a \in A} S_a$	The set of all elements which belong to S for any a of A .
$\cup_{X \subset C} X \ (\cap_{X \subset C} X)$	The join (meet) of the sets $X_i, i = 1, 2, \dots, n$, where each $X_i \subset C$ for a class C .
$x < y$	x is less than y .
$x > y$	x is greater than y .
g.l.b.	Greatest lower bound.
l.u.b.	Least upper bound.
$\sum_{i=1}^n$ or $\sum_i i = 1, 2, \dots, n$	Sum of n terms, one for each positive integer from 1 to n .
$\prod_{i=1}^n$ or $\prod_i i = 1, 2, \dots, n$	Product of n terms, one for each positive integer from 1 to n .
$x \equiv a \pmod{m}$	x is congruent to a modulo m .
g.c.d.	Great common divisor.
l.c.m.	Least common multiple.
(a, b)	The g.c.d. of a and b .
$[a, b]$	The l.c.m. of a and b .
$a \mid b$	a divides b .
$a \nmid b$	a does not divide b .
$a(x), b(x), \dots,$ $f(x), g(x), \dots$	Polynomials in x .
$R[x]$	A set of polynomials in x with coefficients in a ring (as in §4.1.2.5, or the rational number field as in §5.2.1-3) R . R may be replaced by C, D, F, I, \bar{R} , etc.; viz. $C[x], D[x], F[x], I[x], \bar{R}[x]$, denoting the set of polynomials in x with coefficients in C, D, F, I, \bar{R} , respectively.
$\deg f(x)$	The degree of $f(x)$.
$ a_{ij} $	The determinant whose element in the i th row and the j th column is a_{ij} .
(a_{ij}) (or $[a_{ij}]$)	The matrix whose element in the i th row and the j th column is a_{ij} .
A^T	The transpose of a matrix A .
A^*	The adjoint of a matrix A .
A_{ij}	The cofactor of a_{ij} in $A = (a_{ij})$.
\bar{A}_{ij}	An $(n-1)$ by $(n-1)$ submatrix of an n by n matrix $A = (a_{ij})$, i.e. a minor of A .
$\text{Re}(z)$	The real part of a complex number z .
$\text{Im}(z)$	The imaginary part of z .
\bar{z}	The conjugate of z (a complex number, or a Gaussian number, or a Gaussian integer, or an algebraic integer).
$F[a]$	A (simple) algebraic extension of F .
$F[a, b, \dots]$	A multiple algebraic extension of F .
$N(g)$	The norm of g .
$T(g)$	The trace of g .

Mathematical Logic

§1.1.1 Tautologies

Df. 1.1.1.1 *Logic* is analysis of language, which consists of *signs*.

Since signs do not always represent a language, the signs at issue are only some particular signs conventionally coordinated to some significant objects, concrete or abstract. Of such signs, the most fundamental and purposeful signs are propositions.

Df. 1.1.1.2 A *proposition* is an assertive statement (or sentence), which is composed of several words and has a *truth value*, i.e. it can be *true* or *false*.

Example:

“This is white” is a proposition while “May God bless you!” or “Who are you?” is not.

A proposition, then, is not merely a sentence or statement, much less a definitely exclamatory or interrogative (or generally emotional or volitional) statement; it is, as a matter of fact, a cognitive statement which must be verifiable as true or false.

MTh. 1.1.1.3 (Principle of the Identity of Indiscernibles). Two propositions are of the same *meaning* if they cannot be discerned differently for all possible verifications.

Example:

“This is white”, “Dies ist weiss” (in German), and “Ceci est blanc” (in French) are all of the same meaning despite their symbolic differences; so are also the following two propositions in the same language: “Men are two-footed animals” and “Men are bipeds”.

This first metatheorem (i.e. theorem of theorems) is one of the most fundamental of all logical principles, explicitly formulated by Leibniz and called “*principium identitatis indiscernibilium*” (which is in fact a modification of the so-called Occam’s razor: entities should not be multiplied unless necessary). The principle is indeed the core of nominalism which is the backbone of modern mathematics.

Df. 1.1.1.4 Given some already approved propositions, the process of obtaining new propositions solely by virtue of the form and not the content of the original propositions is called *logical* (or *deductive*) *inference*.

Such logical inferences may be symbolized, as in mathematics and mathematical logic, but at the very beginning a great emphasis should be put on the fact, which may be inferred from Gödel’s theorem (which lies beyond the scope and purpose of this book), that a single system of formal logic cannot embrace all forms of reasoning which are correct. Stated otherwise, mathematical reasoning or mathematics in general is but one system of formal logic which as such must suffer from limitations imposed on itself by itself; one of such limitations is, for instance, “implication” (cf. Df. 1.1.1.6, i below).

Df. 1.1.1.5 Propositions may be composite, i.e. made up of *subpropositions* by the following *connectives* (or *logical constants*): negation, disjunction, and conjunction, of which negation is called a *unary* connective and disjunction or conjunction a *binary* connective.

(i) *Negation*, defined by the adjoining table where p denotes a proposition and 1 and 0 represent “true” and “false” respectively, which in turn define \bar{p} which reads “not- p ” and denotes a proposition which is *not* p . Hence, as the table shows, \bar{p} is false if p is true, and true if p is false.

p	\bar{p}
1	0
0	1

Note. If p is negated more than two or three times, the bars above p may be set in front of p ; e.g. $---p$ instead of $\bar{\bar{\bar{p}}}$ (cf. Problem 15, iv).

(ii) *Disjunction*, defined by the table at right, where 1 and 0 denote as above, p and q designate two propositions, and $p \vee q$ reads “ p or q (or both)” and means p or (in the sense of and/or) q . Hence the disjunction as such is the so-called *inclusive* disjunction in contrast with the *exclusive* (or *complete*) disjunction, denoted by $p \vee\vee q$ (cf. Problem 1).

p	q	$p \vee q$
1	1	1
1	0	1
0	1	1
0	0	0

(iii) *Conjunction*, defined by the table at right $p, q, 1, 0$ denoting as above and $p \cdot q$ reading “ p and q ” and designating the same. The dot may be replaced by an upside-down wedge \wedge or may disappear completely, viz. pq , just as for multiplication in elementary algebra. In the latter case the function of parentheses also will be the same as in elementary algebra; e.g. $p(q \vee r) \equiv p \cdot (q \vee r)$ and $pq \vee r \equiv (p \cdot q) \vee r$. This practice will be adopted throughout Part 1.

p	q	$p \cdot q$
1	1	1
1	0	0
0	1	0
0	0	0

These three may be considered the *primary* connectives in the sense that, on the strength of them, two *secondary* connectives may be obtained as follows.

Df. 1.1.1.6

(i) (*Material*) *implication*, defined by the table at right, again $p, q, 1, 0$ denoting the same as above, and $p \rightarrow q$ reading “if p , then q ” (or “ p implies q ” or “ p only if q ”). This connective is redundant, since it can be proved (cf. Problem 9) to be identical with, and may be replaced by, either $\bar{p} \vee q$ or $p\bar{q}$.

p	q	$p \rightarrow q$
1	1	1
1	0	0
0	1	1
0	0	1

(ii) (*Logical*) *equivalence*, defined by the table at right, $p, q, 1, 0$ designating the same as above, and $p \leftrightarrow q$ (or $p \equiv q$) reading “ p if and only if q ” [or “ p is (materially or logically) equivalent to q ”]. This connective is also redundant, since it can be proved (cf. Problem 10) to be indiscernible from, hence may be replaced by, $(p \rightarrow q)(q \rightarrow p)$, i.e. $(\bar{p} \vee q)(\bar{q} \vee p)$ or $(p\bar{q})(q\bar{p})$.

p	q	$p \leftrightarrow q$
1	1	1
1	0	0
0	1	0
0	0	1

It must be emphasized that the “if-then” defined by (material) implication is somewhat different from what is meant by “if” and “then” in everyday language, mainly because the ordinary “if-then” often designates causal relations, which are more physical than logical. The implication in mathematical logic is to mean neither more nor less than “not- p or q ” or “it is not the case that p and not- q ”.

Example:

“ p ” and “ q ” representing “two lines are parallel” and “two lines do not intersect” (in Euclidean space) respectively, “ $p \rightarrow q$ ” denotes “if two lines are parallel, then the two lines do not intersect”,

whose meaning in mathematical logic is identical with “it is not the case that two lines are parallel and it is not the case that the two lines do not intersect”, i.e. “it is false that two lines are parallel and they intersect”.

Likewise, “ p ” and “ q ” representing “two triangles T_1 and T_2 are similar” and “the corresponding sides of T_1 and T_2 are proportional” respectively, “ $p \leftrightarrow q$ ” denotes “if T_1 and T_2 are similar, then the corresponding sides of T_1 and T_2 are proportional, and if the corresponding sides of T_1 and T_2 are proportional, then T_1 and T_2 are similar” or “if T_1 and T_2 are similar, then the corresponding sides of T_1 and T_2 are proportional, and conversely” or “ T_1 and T_2 are similar if and only if the corresponding sides of T_1 and T_2 are proportional” or “the corresponding sides of T_1 and T_2 are proportional if and only if T_1 and T_2 are similar”.

Notice the difference in the meaning of “if and only if” exemplified above and “if and only if” in everyday language. If this example is to be interpreted in everyday language, it becomes immediately false or at best inadequate, since “ T_1 and T_2 are similar” holds *also* when “the corresponding angles of T_1 and T_2 are equal”. Mathematical language is, to repeat, not identical with everyday language.

Note. “if and only if” will be abbreviated as “iff” throughout this book.

Df. 1.1.1.7 A *tautology* is a proposition which is true for all truth-values of its sub-propositions.

Example:

The proposition: $p \rightarrow q \equiv \bar{p} \vee q$ (or $p \rightarrow q \equiv \bar{p}\bar{q}$ or $\bar{p} \vee q \equiv \bar{p}\bar{q}$) is a tautology, since its truth-value as a whole is always 1 for every possible choice of truth-values for p and q . (Cf. Prob. 10.)

Df. 1.1.1.8 Any negated tautology, which therefore must always be false, is called a *contradiction*.

Example:

The negation of a tautology: $p \rightarrow p$ (cf. Prob. 15, i below) is $\overline{p \rightarrow p}$, and $\overline{p \rightarrow p} \equiv \bar{p} \vee p$ since $p \rightarrow p \equiv \bar{p} \vee p$ by Df. 1.1.1.6, i. Hence, by breaking negation lines (cf. Prob. 12, below), $\overline{p \rightarrow p} \equiv \bar{p} \vee p \equiv \bar{p}\bar{p} \equiv p\bar{p}$ ($\because \bar{p} \equiv p$, cf. Prob. 4 below), and $p\bar{p}$, which reads “ p and not p ” (at the same time) is certainly a contradiction in every sense of the word.

To carry out logical inferences the following principles must be first taken for granted.

MTh. 1.1.1.9 (Principle of Substitution). Proper *substitutions* do not affect the truth-value of tautologies.

Proper substitutions consist of either substitutions on *variables*, i.e. the symbols which denote propositions, or definitional substitutions. E.g. if $p \rightarrow q \equiv \bar{p} \vee q$ is a tautology, it remains a tautology through the substitution of new variables, say, \bar{a} and b in the place of p and q respectively; i.e. $\bar{a} \rightarrow b \equiv \bar{\bar{a}} \vee b$ is a tautology just as its counterpart in terms of p and q is a tautology and as long as the substitution is carried out completely and consistently. (Hence $\bar{a} \rightarrow b \equiv \bar{p} \vee q$, for instance, is not *ipso facto* a tautology unless, of course, there are some additional stipulations.)

Likewise a definition itself may serve as a substitution if one definition is logically equivalent to the other; e.g. $p\bar{q} \vee \bar{p}q$ may be replaced by $p \vee q$ whenever and wherever it is convenient to do so once the former is defined [or, in this particular case (cf. Prob. 1), proved] to be the same as the latter.

The fundamental principle of substitution is followed by a group of metatheorems (which may be classified in many ways, depending on the taste of authors).

MTh. 1.1.1.10 If a proposition q is deductible by MTh. 1.1.1.9 from p , which may be a tautologous proposition or a set of tautologous propositions, then " $p \rightarrow q$ " is a tautology.

Example:

The well-known five tautologies of the *Principia Mathematica* by Whitehead-Russell constitute such a set, which runs as follows:

- P1:** Principle of Tautology. $a \vee a \rightarrow a$.
P2: Principle of Addition. $a \rightarrow a \vee b$.
P3: Principle of Permutation. $a \vee b \rightarrow b \vee a$.
P4: Principle of Summation. $(b \rightarrow c) \rightarrow (a \vee b \rightarrow a \vee c)$.
P5: Principle of Association. $a \vee (b \vee c) \rightarrow b \vee (a \vee c)$.

Note. Such tautologies, called the *primitives* (or *postulates* or *axioms*), must be consistent and complete, as **P1-5** are, but may not be independent, as **P1-5** are not; e.g. **P5** is deducible from the rest. (cf. Prob. 17 below).

MTh. 1.1.1.11 If " $p \rightarrow q$ " is true and " p " is true, then " q " is true.

Example:

" p " and " q " representing "an infinite series converges" and "the general term of the given series approaches zero" respectively, the logical inference of this metatheorem takes the following form:

- (i) " $p \rightarrow q$ " is true: "if an infinite series converges, then the general term of the given series approaches zero" (which is a true theorem of the Calculus).
(ii) " p " is true: "an infinite series converges".
 \therefore " q " is true: "the general term of the given series approaches zero" [which is true if (i) and (ii) are true].

This rule is often called the Principle of Inference or *modus ponens*, a name inherited from medieval logic.

MTh. 1.1.1.12 If " $p \rightarrow q$ " is true and " q " is false, then " p " is false.

Example:

" p " and " q " representing "a function $f(x)$ is differentiable at $x = x_0$ " and " $f(x)$ is continuous at $x = x_0$ " respectively, this metatheorem is the logical inference of the following form:

- (i) " $p \rightarrow q$ " is true: "if a function $f(x)$ is differentiable at $x = x_0$, then $f(x)$ is continuous at $x = x_0$ " (which is a true theorem of the Calculus).
(ii) " q " is false: " $f(x)$ is continuous at $x = x_0$ " is false, i.e. " $f(x)$ is discontinuous at $x = x_0$ " is true.
 \therefore " p " is false: " $f(x)$ is differentiable at $x = x_0$ " is false, i.e. " $f(x)$ cannot be differentiated at $x = x_0$ " is true [which is true if (i) and (ii) hold].

Stated otherwise: if " $p \rightarrow q$ " is true and " \bar{q} " is true, then " \bar{p} " is also true; or, stated more differently: if " $p \rightarrow q$ " is true, then " $\bar{q} \rightarrow \bar{p}$ " is also true (cf. Prob. 13).

This rule also has a medieval name, *modus tollens*, or it is called the Principle of Negative Inference (or Contraposition).

MTh. 1.1.1.13 If " $p \rightarrow q$ " is true and " $q \rightarrow r$ " is true, then " $p \rightarrow r$ " is true.

Example:

" p ", " q ", and " r " designating "a function $f(x)$ is differentiable at $x = x_0$ ", " $f(x)$ is continuous at $x = x_0$ ", and " $f(x)$ is integrable at $x = x_0$ " respectively, the logical pattern of this metatheorem runs as follows:

- (i) “ $p \rightarrow q$ ” is true: “if a function $f(x)$ is differentiable at $x = x_0$, then $f(x)$ is continuous at $x = x_0$ ” is true (which is in fact true).
- (ii) “ $q \rightarrow r$ ” is true: “if $f(x)$ is continuous at $x = x_0$, then $f(x)$ is integrable at $x = x_0$ ” is true (which is also proved to be true).
- \therefore “ $p \rightarrow r$ ” is true: “if $f(x)$ is differentiable at $x = x_0$, then $f(x)$ is integrable at $x = x_0$ ” is true (which is logically true).

Generalized, this metatheorem has the following form:

$$“p_1 \rightarrow p_2”, “p_2 \rightarrow p_3”, \dots, \text{ and } “p_{n-1} \rightarrow p_n” \text{ imply } “p_1 \rightarrow p_n”.$$

In this sense it has a descriptive name: Chain Rule (or Syllogism Principle, as it is called in the *Principia Mathematica*).

MTh. 1.1.1.14 If “ p ” is true and “ q ” is true, then “ pq ” is true.

Example:

“ p ” and “ q ” denoting “a number n is an integer” and “ n is positive” (in the same context) respectively, this metatheorem has the following scheme:

- (i) “ p ” is true: “a number n is an integer” is true.
- (ii) “ q ” is true: “ n is positive” is true (in the same context).
- \therefore “ pq ” is true: “a number n is an integer and it is positive” is true, i.e. “ n is a positive integer” is true (in the given context).

This rule is called the Principle of Adjunction.

MTh. 1.1.1.15 There exist two rules of disjunctive inference:

- (i) *Modus tollendo ponens*: if “ $p \vee q$ ” is true and “ p ” is false, then “ q ” is true.
- (ii) *Modus ponendo tollens*: if “ $p \vee q$ ” is true and “ p ” is true, then “ q ” is false.

The validity of this metatheorem can be readily exemplified by letting, for instance, “ p ” and “ q ” represent “a number x is an integer” and “ x is a real number” respectively for (i) and “a number n is odd” and “ n is even” respectively for (ii).

MTh. 1.1.1.16 There exists an equivalence inference: if “ $p \equiv q$ ” is true and “ p ” is true, then “ q ” is true.

Example:

“ p ” and “ q ” representing “two triangles T_1 and T_2 are similar” and “ T_1 and T_2 are congruent” respectively, it is evident that “ T_1 and T_2 are congruent” is true if “two triangles T_1 and T_2 are similar iff T_1 and T_2 are congruent” is true and “ T_1 and T_2 are similar” is true.

Solved Problems

1. Analyse the concept of exclusive (or complete) disjunction in terms of connectives, then verify it by a truth table.

PROOF:

(i) Since the exclusive disjunction is defined by “ p or q but not both”, it can be true when and only when one and only one of p or q is true. Stated otherwise: “ p or q but not both” must be identical with “ p and not- q or not- p and q ” or “ p or q and it is not the case that both p and q hold”; i.e. if “ \vee ” is to denote the exclusive “or”, then it must be proved to be a tautology that

$$p \vee q \equiv p\bar{q} \vee \bar{p}q \quad \text{or} \quad p \vee q \equiv (p \vee q)(\bar{p}\bar{q})$$

(ii) The tautologies are demonstrated as follows:

1		2		3		4	5	6
p	q	\bar{p}	\bar{q}	$p\bar{q}$	$\bar{p}q$	$p\bar{q} \vee \bar{p}q$	$p \vee q$	$p \vee q \equiv p\bar{q} \vee \bar{p}q$
1	1	0	0	0	0	0	0	1
1	0	0	1	1	0	1	1	1
0	1	1	0	0	1	1	1	1
0	0	1	1	0	0	0	0	1

The truth-table above is numbered to show that the demonstration consists of six steps. Step 1 is justified by the fact, as in Df. 1.1.1.5-6, that there cannot be other alternatives (i.e. in the two-value logic of "true" and "false") for two propositions p and q . Step 2 follows from Step 1, by Df. 1.1.1.5, i. Step 3 is obtained by Step 2 and Df. 1.1.1.5, iii. Step 4 follows from Step 3 and Df. 1.1.1.5, ii. Step 5 is the result of the original analysis of the concept itself. Finally, Step 6 is obtained from Steps 4, 5 and Df. 1.1.1.6, ii. Since Step 6 shows that the proposition is true on all occasions, i.e. a tautology, the proof is complete.

$p \vee q \equiv (p \vee q)(\bar{p}\bar{q})$ can be proved likewise.

Note. " $p \vee q$ " is sometimes considered an exclusive and complete disjunction — exclusive, because at most one term of the disjunction is true, and complete, because at least one of the terms is true, i.e. the disjunction is true.

2. Show that " $p|q$ ", which reads " p and q are not both true", symbolized by the stroke " $|$ ", called the *alternative denial* and defined by the truth-table at right, makes all the primary connectives of Df. 1.1.1.5 deducible from itself.

p	q	$p q$
1	1	0
1	0	1
0	1	1
0	0	1

PROOF:

The three primary connectives may be expressed in terms of strokes, defined as above, as follows:

$$(i) \bar{p} \equiv p|p, \quad (ii) p \vee q \equiv (p|p) | (q|q), \quad (iii) pq \equiv (p|q) | (p|q)$$

each of which is a tautology, as can be readily verified by a truth-table, e.g. with respect to (ii):

p	q	$p \vee q$	$p p$	$q q$	$(p p) (q q)$	$p \vee q \equiv (p p) (q q)$
1	1	1	0	0	1	1
1	0	1	0	1	1	1
0	1	1	1	0	1	1
0	0	0	1	1	0	1

(i) and (iii) can be proved likewise.

3. Prove that " $p \downarrow q$ ", which reads "neither p nor q is true" (i.e. p and q are both false), symbolized by the dagger " \downarrow ", called the *joint denial* and defined by the truth table at right, works exactly the same way as the alternative denial with respect to the primary connectives of Df. 1.1.1.5; i.e. they may be replaced by the joint denial.

p	q	$p \downarrow q$
1	1	0
1	0	0
0	1	0
0	0	1

PROOF:

The three primary connectives may be expressed in terms of daggers, defined as above, as follows:

$$(i) \bar{p} \equiv p \downarrow p, \quad (ii) p \vee q \equiv (p \downarrow q) \downarrow (p \downarrow q), \quad (iii) pq \equiv (p \downarrow p) \downarrow (q \downarrow q)$$

each of which is a tautology, as can be verified by a truth-table as in Prob. 2.

Note. The truth-table above begins with three columns for three initial subpropositions, constituting a *ternary* matrix of propositions in contrast with the preceding *unary* (or *monary*) matrices (cf. Df. 1.1.1.5, i; Prob. 4, 6) and *binary* matrices (cf. Df. 1.1.1.5, ii, iii; Prob. 1, etc.). In this sense there exist *quaternary* matrices of propositions (cf. Prob. 8 below) or *quinary* or, in general, *n-ary* matrices of propositions, depending on the number of initial subpropositions. The number of the rows of truth-tables, then, will grow with the number of initial subpropositions; e.g. a septenary matrix of propositions has $2^7 = 128$ rows and, in general, a *n-ary* matrix has 2^n rows, which may be so many as to incapacitate manual truth-table computations.

8. Prove that implications may merge as follows:

- (i) $(p \rightarrow q) \vee (p \rightarrow r) \equiv p \rightarrow q \vee r$ (iv) $(p \rightarrow r)(q \rightarrow r) \equiv p \vee q \rightarrow r$
(ii) $(p \rightarrow q)(p \rightarrow r) \equiv p \rightarrow qr$ (v) $(p \rightarrow q)(r \rightarrow s) \rightarrow (pr \rightarrow qs)$
(iii) $(p \rightarrow r) \vee (q \rightarrow r) \equiv pq \rightarrow r$ (vi) $(p \rightarrow q)(r \rightarrow s) \rightarrow (p \vee r \rightarrow q \vee s)$

PROOF:

Since the six propositions have similar proofs, (i) and (vi) are considered their representatives.

<i>p</i>	<i>q</i>	<i>r</i>	$p \rightarrow q$	$p \rightarrow r$	$q \vee r$	$(p \rightarrow q) \vee (p \rightarrow r)$	$p \rightarrow q \vee r$	$(p \rightarrow q) \vee (p \rightarrow r) \equiv p \rightarrow q \vee r$
1	1	1	1	1	1	1	1	1
1	1	0	1	0	1	1	1	1
1	0	1	0	1	1	1	1	1
1	0	0	0	0	0	0	0	1
0	1	1	1	1	1	1	1	1
0	1	0	1	1	1	1	1	1
0	0	1	1	1	1	1	1	1
0	0	0	1	1	0	1	1	1

<i>p</i>	<i>q</i>	<i>r</i>	<i>s</i>	$p \rightarrow q$	$r \rightarrow s$	$p \vee r$	$q \vee s$	$(p \rightarrow q)(r \rightarrow s)$	$p \vee r \rightarrow q \vee s$	$(p \rightarrow q)(r \rightarrow s) \rightarrow (p \vee r \rightarrow q \vee s)$
1	1	1	1	1	1	1	1	1	1	1
1	1	1	0	1	0	1	1	0	1	1
1	1	0	1	1	1	1	1	1	1	1
1	1	0	0	1	1	1	1	1	1	1
1	0	1	1	0	1	1	1	0	1	1
1	0	1	0	0	0	1	0	0	0	1
1	0	0	1	0	1	1	1	0	1	1
1	0	0	0	0	1	1	0	0	0	1
0	1	1	1	1	1	1	1	1	1	1
0	1	1	0	1	0	1	1	0	1	1
0	1	0	1	1	1	0	1	1	1	1
0	1	0	0	1	1	0	1	1	1	1
0	0	1	1	1	1	1	1	1	1	1
0	0	1	0	1	0	1	0	0	0	1
0	0	0	1	1	1	0	1	1	1	1
0	0	0	0	1	1	0	0	1	1	1

9. Implications may be dissolved as follows: (i) $p \rightarrow q \equiv \bar{p} \vee q$, (ii) $p \rightarrow q \equiv p\bar{q}$.

PROOF:

(i)

p	q	\bar{p}	$\bar{p} \vee q$	$p \rightarrow q$	$p \rightarrow q \equiv \bar{p} \vee q$
1	1	0	1	1	1
1	0	0	0	0	1
0	1	1	1	1	1
0	0	1	1	1	1

(ii) can be proved likewise.

10. $p \equiv q$ iff $(p \rightarrow q)(q \rightarrow p)$; i.e. $(p \equiv q) \equiv (p \rightarrow q)(q \rightarrow p)$ is a tautology, and so is $(p \equiv q) \equiv (\bar{p} \vee q)(\bar{q} \vee p)$ or $(p \equiv q) \equiv (p\bar{q})(q\bar{p})$.

PROOF:

Problem 9 has already proved that $p \rightarrow q \equiv \bar{p} \vee q$ and $p \rightarrow q \equiv p\bar{q}$ and that, likewise, $q \rightarrow p \equiv \bar{q} \vee p$ and $q \rightarrow p \equiv q\bar{p}$. Hence the proof is complete if a truth-table justifies the first part of the problem, viz.:

p	q	$p \rightarrow q$	$q \rightarrow p$	$(p \rightarrow q)(q \rightarrow p)$	$p \equiv q$	$(p \equiv q) \equiv (p \rightarrow q)(q \rightarrow p)$
1	1	1	1	1	1	1
1	0	0	1	0	0	1
0	1	1	0	0	0	1
0	0	1	1	0	1	1

11. Both implications and equivalences are transitive, i.e.,

$$(i) \quad (p \rightarrow q)(q \rightarrow r) \rightarrow (p \rightarrow r), \quad (ii) \quad (p \equiv q)(q \equiv r) \rightarrow (p \equiv r)$$

PROOF:

Since both proofs are similar, only (ii) is proved:

p	q	r	$p \equiv q$	$q \equiv r$	$(p \equiv q)(q \equiv r)$	$p \equiv r$	$(p \equiv q)(q \equiv r) \rightarrow (p \equiv r)$
1	1	1	1	1	1	1	1
1	1	0	1	0	0	0	1
1	0	1	0	0	1	1	1
1	0	0	0	1	0	0	1
0	1	1	0	1	0	0	1
0	1	0	0	0	1	1	1
0	0	1	1	0	0	0	1
0	0	0	1	1	1	1	1

12. Negation lines may be broken as follows:

$$(i) \quad \overline{p \vee q} \equiv \bar{p} \cdot \bar{q}, \quad (ii) \quad \overline{pq} \equiv \bar{p} \vee \bar{q}, \quad (iii) \quad \overline{p \equiv q} \equiv (\bar{p} \equiv q) \equiv (p \equiv \bar{q})$$

PROOF:

Because of similarity, the proof of (iii) alone is shown:

p	q	r	\bar{p}	\bar{q}	$p \equiv q$	$\overline{p \equiv q}$	$\bar{p} \equiv q$	$p \equiv \bar{q}$	$\overline{p \equiv q} \equiv (\bar{p} \equiv q)$	$(\bar{p} \equiv q) \equiv (p \equiv \bar{q})$
1	1	1	0	0	1	0	0	0	1	1
1	1	0	0	0	1	0	0	0	1	1
1	0	1	0	1	0	1	1	1	1	1
1	0	0	0	1	0	1	1	1	1	1
0	1	1	1	0	0	1	1	1	1	1
0	1	0	1	0	0	1	1	1	1	1
0	0	1	1	1	1	0	0	0	1	1
0	0	0	1	1	1	0	0	0	1	1

And, by Prob. 10, $\overline{p \equiv q} \equiv (p \equiv \bar{q})$; hence $\overline{p \equiv q} \equiv (\bar{p} \equiv q) \equiv (p \equiv \bar{q})$.

(Or, by observation, all three have exactly the same truth values, justifying the conclusion.)

13. $\bar{q} \rightarrow \bar{p}$ iff $p \rightarrow q$; i.e. $p \rightarrow q \equiv \bar{q} \rightarrow \bar{p}$ is a tautology.

PROOF:

p	q	\bar{p}	\bar{q}	$p \rightarrow q$	$\bar{q} \rightarrow \bar{p}$	$p \rightarrow q \equiv \bar{q} \rightarrow \bar{p}$
1	1	0	0	1	1	1
1	0	0	1	0	0	1
0	1	1	0	1	1	1
0	0	1	1	1	1	1

Note. $\bar{q} \rightarrow \bar{p}$ is called the *contrapositive* (or *opposite converse*) of $p \rightarrow q$.

14. An arbitrary term or factor or implication itself may be added to implications as follows:

$$(i) \quad p \rightarrow p \vee q, \quad (ii) \quad (p \rightarrow q) \rightarrow (p \rightarrow q \vee r), \quad (iii) \quad (p \rightarrow q) \rightarrow (pr \rightarrow q), \\ (iv) \quad p \rightarrow (q \rightarrow p), \quad (v) \quad \bar{p} \rightarrow (p \rightarrow q)$$

On the other hand, a term or factor in implications may be dropped as follows:

$$(vi) \quad (p \vee r \rightarrow q) \rightarrow (p \rightarrow q), \quad (vii) \quad (p \rightarrow qr) \rightarrow (p \rightarrow q)$$

and any term or factor proved to be always true or false may be dropped as follows:

$$(viii) \quad p(q \vee \bar{q}) \equiv p, \quad (ix) \quad p \vee q\bar{q} \equiv p$$

PROOF:

All nine propositions are tautologies whose proofs are quite readily verifiable by simple truth tables.

15. Deduce the following propositions solely by MTh. 1.1.1.9-16:

$$(i) \quad p \rightarrow p, \quad (ii) \quad p \vee \bar{p}, \quad (iii) \quad p \rightarrow \bar{\bar{p}}, \quad (iv) \quad \bar{\bar{p}} \rightarrow p$$

PROOF:

(i)	$p \rightarrow p \vee p$	by MTh. 1.1.1.10, P2
	$p \vee p \rightarrow p$	by MTh. 1.1.1.10, P1
\therefore	$p \rightarrow p$	by MTh. 1.1.1.13
(ii)	$p \rightarrow p$	by (i) above
	$\bar{p} \vee p$	Df. by MTh. 1.1.1.9
	$\bar{p} \vee p \rightarrow p \vee \bar{p}$	by MTh. 1.1.1.10, P3
\therefore	$p \vee \bar{p}$	by MTh. 1.1.1.11
(iii)	$\bar{p} \vee \bar{\bar{p}}$	by (ii) and MTh. 1.1.1.9
\therefore	$p \rightarrow \bar{\bar{p}}$	Df. ($p \rightarrow q \equiv \bar{p} \vee q$) by MTh. 1.1.1.9
(iv)	$\bar{p} \rightarrow \text{---}p$	by (iii) above
	$(\bar{p} \rightarrow \text{---}p) \rightarrow [(p \vee \bar{p}) \rightarrow (\bar{p} \vee \text{---}p)]$	MTh. 1.1.1.10, P4
	$(p \vee \bar{p}) \rightarrow (p \vee \text{---}p)$	MTh. 1.1.1.11
	$p \vee \bar{p}$	by (ii) above
	$p \vee \text{---}p$	MTh. 1.1.1.11
	$(p \vee \text{---}p) \rightarrow (\text{---}p \vee p)$	MTh. 1.1.1.10, P3
	$\text{---}p \vee p$	MTh. 1.1.1.11
\therefore	$\bar{\bar{p}} \rightarrow p$	Df., as in (iii), by MTh. 1.1.1.9

16. Deduce, as in Prob. 15, the following propositions:

$$(i) \quad p \rightarrow pp, \quad (ii) \quad pq \rightarrow p, \quad (iii) \quad (p \rightarrow q) \rightarrow (\bar{q} \rightarrow \bar{p})$$

PROOF:

(i)	$(\bar{p} \vee \bar{p}) \rightarrow \bar{p}$	MTh. 1.1.1.10, P1
	$\bar{\bar{p}} \rightarrow (\bar{p} \vee \bar{p})$	MTh. 1.1.1.12
	$p \rightarrow \bar{\bar{p}}$	Prob. 15, iii
	$p \rightarrow \bar{p} \vee \bar{p}$	MTh. 1.1.1.13
\therefore	$p \rightarrow pp$	Df. ($pq \equiv \bar{\bar{p}} \vee \bar{q}$, cf. Prob. 12, i) by MTh. 1.1.1.9
(ii)	$\bar{p} \rightarrow \bar{p} \vee \bar{q}$	MTh. 1.1.1.10, P2
	$\bar{\bar{p}} \vee \bar{q} \rightarrow \bar{\bar{p}}$	MTh. 1.1.1.12
	$\bar{\bar{p}} \rightarrow p$	Prob. 15, iv
	$\bar{\bar{p}} \vee \bar{q} \rightarrow p$	MTh. 1.1.1.13
\therefore	$pq \rightarrow p$	Df. [cf. (i) above] by MTh. 1.1.1.9
(iii)	$(q \rightarrow \bar{\bar{q}}) \rightarrow [(\bar{p} \vee q) \rightarrow (\bar{p} \vee \bar{\bar{q}})]$	MTh. 1.1.1.10, P4
	$q \rightarrow \bar{\bar{q}}$	Prob. 15, iii
	$\bar{p} \vee q \rightarrow \bar{p} \vee \bar{\bar{q}}$	MTh. 1.1.1.11
	$\bar{p} \vee \bar{\bar{q}} \rightarrow \bar{\bar{q}} \vee \bar{p}$	MTh. 1.1.1.10, P3
	$\bar{p} \vee q \rightarrow \bar{\bar{q}} \vee \bar{p}$	MTh. 1.1.1.13
\therefore	$(p \rightarrow q) \rightarrow (\bar{q} \rightarrow \bar{p})$	Df. ($\bar{p} \vee q \equiv p \rightarrow q$) by MTh. 1.1.1.9

17. Prove the redundancy of **P5** in MTh.1.10 by deducing it from **P1-4**.

PROOF:

$r \rightarrow r \vee p$	P2
$r \vee p \rightarrow p \vee r$	P3
$r \rightarrow p \vee r$	MTh. 1.1.1.13
$(r \rightarrow p \vee r) \rightarrow [q \vee r \rightarrow q \vee (p \vee r)]$	P4
$q \vee r \rightarrow q \vee (p \vee r)$	MTh. 1.1.1.11
$[q \vee r \rightarrow q \vee (p \vee r)] \rightarrow \{p \vee (q \vee r) \rightarrow p \vee [q \vee (p \vee r)]\}$	P4
$p \vee (q \vee r) \rightarrow p \vee [q \vee (p \vee r)]$	MTh. 1.1.1.11
$p \vee [q \vee (p \vee r)] \rightarrow [q \vee (p \vee r)] \vee p$	P3
$p \vee (q \vee r) \rightarrow [q \vee (p \vee r)] \vee p$	MTh. 1.1.1.13
$p \vee r \rightarrow (p \vee r) \vee q$	P2
$(p \vee r) \vee q \rightarrow q \vee (p \vee r)$	P3
$p \vee r \rightarrow q \vee (p \vee r)$	MTh. 1.1.1.13
$p \rightarrow p \vee r$	P2
$[p \rightarrow q \vee (p \vee r)] \rightarrow \{[q \vee (p \vee r)] \vee p \rightarrow \{[q \vee (p \vee r)] \vee [q \vee (p \vee r)]\}\}$	P4
$[q \vee (p \vee r)] \vee p \rightarrow \{[q \vee (p \vee r)] \vee [q \vee (p \vee r)]\}$	MTh. 1.1.1.11
$[q \vee (p \vee r)] \vee [q \vee (p \vee r)] \rightarrow [q \vee (p \vee r)]$	P1
$[q \vee (p \vee r)] \vee p \rightarrow q \vee (p \vee r)$	MTh. 1.1.1.13

Hence it follows from the last step and the ninth that

$$p \vee (q \vee r) \rightarrow q \vee (p \vee r) \quad \text{MTh. 1.1.1.13}$$

18. Does “ $p \rightarrow \bar{s}$ ” follow from four hypotheses: “ ps ”, “ $p \rightarrow q \vee r$ ”, “ $s \rightarrow \bar{r}$ ”, “ $q \rightarrow \bar{p}$ ”?

PROOF:

It does, since

(1) ps	Hyp ₁
(2) $p \rightarrow q \vee r$	Hyp ₂
(3) $s \rightarrow \bar{r}$	Hyp ₃
(4) $q \rightarrow \bar{p}$	Hyp ₄
(5) $q \vee r$	(1), (2), and MTh. 1.1.1.11
(6) \bar{r}	(1), (3), and MTh. 1.1.1.11
(7) q	(5), (6), and MTh. 1.1.1.15, i
(8) \bar{p}	(4), (7), and MTh. 1.1.1.11
(9) $p\bar{p}$	(1), (8), and MTh. 1.1.1.14
(10) $ps \rightarrow p\bar{p}$	(1)-(9), and MTh. 1.1.1.13
(11) $p\bar{p} \rightarrow \bar{ps}$	(10), and MTh. 1.1.1.12
(12) \bar{ps}	(9), (11), and MTh. 1.1.1.11
(13) $p \rightarrow \bar{s}$	(12), and Df. ($\bar{a}\bar{b} \equiv \bar{a} \vee \bar{b}$ and $a \rightarrow b \equiv \bar{a} \vee b$) by MTh. 1.1.1.9

*§1.1.2 Quantifications

Df. 1.1.2.1 A sign which represents a proposition is called a (*propositional*) *variable*, in contrast with which all connectives, defined by Df. 1.1.1.5-6, are called *constants*.

Example:

p, q, r , etc. throughout §1.1.1 are all variables, where each variable preserves a recognizable identity in various occurrences for a definite context.

Df. 1.1.2.2 Any combination of concepts which involves one or more variables is a (*propositional*) *function*, which becomes a proposition whenever its variables take values and become specified.

Example:

" x is a real number" is a combination of concepts which contains a variable x and as such is a propositional function, being neither true nor false; it takes value and becomes a proposition iff it is specified, e.g. $x = \sqrt{2}$. Note that a proposition like " $p \vee q$ " or " $p \rightarrow q$ " is actually a propositional function as long as no specified values are assigned to both p and q .

Df. 1.1.2.3 A propositional function f of one variable x , denoted by $f(x)$, may be satisfied by *all* values of x or *some* value or values of x or *no* values of x . The first case is denoted by $(x)f(x)$, which reads "for all values of x , $f(x)$ is true", and (x) is called a *universal quantifier*. The second case is denoted by $(Ex)f(x)$, which reads "there exists a value of x such that $f(x)$ is true", and (Ex) is called an *existential quantifier*. The proposition $(x)f(x)$ in entirety is called a *universal proposition*, and the proposition $(Ex)f(x)$ an *existential proposition*.

Example:

" x is an equilateral triangle", which may be denoted by $L(x)$, is a propositional function; so is " x is an equiangular triangle", denoted by $A(x)$, but their compound "if x is an equilateral triangle, then x is an equiangular triangle" is a universal proposition, since the proposition is valid for *any* x in this specific context. Hence $(x)[L(x) \rightarrow A(x)]$, which reads "for any x , if x is an equilateral triangle, then x is an equiangular triangle".

Note. Different notations are also available for quantified propositions; e.g. $(x)f_x$ or $\forall x f(x)$ or $\{x: f(x)\}$ (or $\{x | f(x)\}$) instead of $(x)f(x)$, and $(\exists x)f(x)$ or $\exists x f_x$ instead of $(Ex)f(x)$. In particular, the form $\{x: \}$ or $\{x | \}$ will sometimes be used in Part 2.

Df. 1.1.2.4 A propositional function is said to lie within the *scope* of the quantifier (x) or (Ex) either if it lies directly to the right of the quantifier or if it is a component of some compound propositional function in parentheses (or brackets or braces) immediately to the right of the quantifier. The variable of the propositional function within the scope is called a *bound* variable and, if otherwise, a *free* variable.

Example:

In the example above: $(x)[L(x) \rightarrow A(x)]$, x of both $L(x)$ and $A(x)$ is bound, while x of $C(x)$ in a context $(Ex)[A(x) \vee B(x)] \rightarrow C(x)$ is free.

MTh. 1.1.2.5 The negation of quantifications is defined as follows:

$$(i) \quad \overline{(x)f(x)} \equiv (Ex)\overline{f(x)} \qquad (ii) \quad \overline{(Ex)f(x)} \equiv (x)\overline{f(x)}$$

and, as it immediately follows,

$$(i') \quad (x)f(x) \equiv \overline{(Ex)\overline{f(x)}} \qquad (ii') \quad (Ex)f(x) \equiv \overline{(x)\overline{f(x)}}$$

The so-called square of opposition from classical logic illustrates the relation among them:

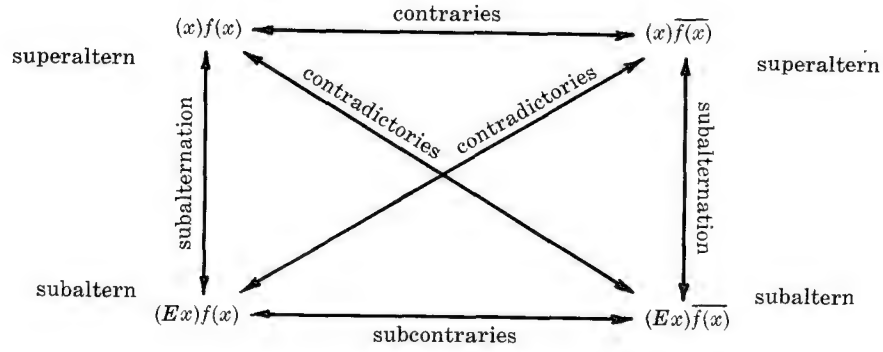


Fig. 1.1.2a

In traditional logic $(x)f(x)$ and $(Ex)f(x)$ are represented by A and I (from *affirmo*) respectively; likewise $(x)\bar{f}(x)$ and $(Ex)\bar{f}(x)$ by E and O (from *nego*) respectively. If S and P represent subject and predicate respectively, then, again in classical logic, SAP designates “all S is P ”, i.e. $(x)f(x)$; SEP “No S is P ”, i.e. $(x)\bar{f}(x)$; SIP “some S is P ”, i.e. $(Ex)f(x)$; and SOP “some S is not P ”, i.e. $(Ex)\bar{f}(x)$.

Note that the third case of Df. 1.1.3.1, i.e. a propositional function satisfied by no values of x , is now represented by SEP , i.e. $(Ex)\bar{f}(x) \equiv (x)\bar{f}(x)$.

Furthermore, the so-called four categorical propositions of A, E, I, O have pictorial representations, called *Venn diagrams*, by drawing two intersecting circles as follows:

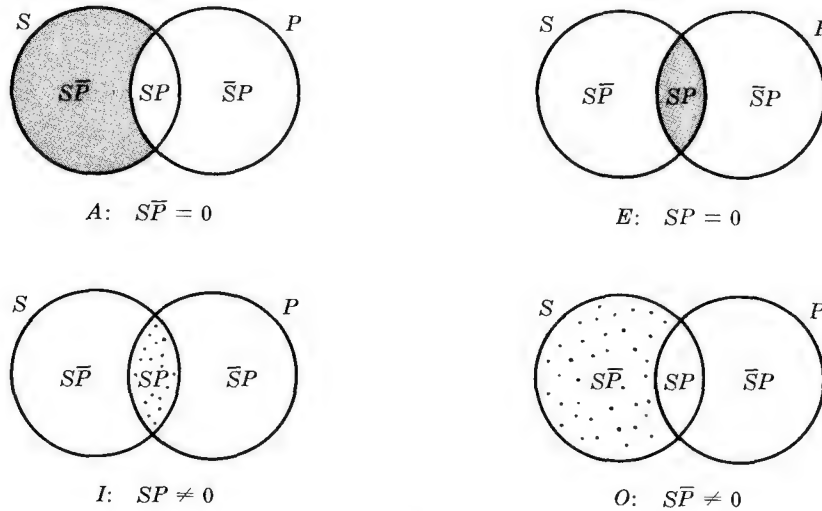


Fig. 1.1.2b

Although the diagrams look self-explanatory, their exact interpretation presupposes a certain amount of knowledge on classes or sets; further explanations, therefore, will be given in Chapter 2.3.

To carry out inferences through quantified propositions, a few new metatheorems must be added to MTh. 1.1.2.5 and the metatheorems of §1.1.1, which are justifiably presumed to remain valid even after going through quantifications. (A detailed, and rather delicate, examination of this presumption lies again beyond the scope of this book.)

MTh. 1.1.2.6 (Principle of Generalization).

- (i) U.G. (Universal Generalization): $(x)f(x)$ is true if $f(x)$ is satisfied by any arbitrary (but significant) values of x .
- (ii) E.G. (Existential Generalization): $(Ex)f(x)$ is true if there exists at least one instance a such that $f(a)$ is true.

Example:

The procedure in geometry of starting with “Let ABC be any triangle”, proving that ABC has a certain property, and ending it with a conclusion that all triangles have the property is a typical case of U.G. On the other hand, solving algebraic or elementary transcendental equations is a familiar case of E.G.: e.g. there exist two roots, real or imaginary, for the equation $ax^2 + bx + c = 0$ where a, b, c are real and $a \neq 0$; or, from trigonometry, there exists a certain set of values which satisfy x in $\sin x + \cos x = 1$; or, from logarithms, there exists a value of x which satisfies the equation $10^x = 3$.

MTh. 1.1.2.7 (Principle of Specialization).

- (i) U.S. (Universal Specialization): $f(a)$ is true for any significant value a for x if $(x)f(x)$ is true.
- (ii) E.S. (Existential Specialization): there exists at least one significant value a for x such that $f(a)$ is true if $(Ex)f(x)$ is true.

Example:

The time-honored syllogism: “All men are mortal; Socrates is a man; therefore Socrates is mortal” is a case of U.S. On the other hand, e.g., “there exist some real numbers which are not rational” justifies “ $\sqrt{2}$ is an irrational number”, exemplifying E.S.

It must be made quite clear at this juncture that, in the process of inference, a free variable (sometimes called a *flagged* variable) in a hypothesis or in any line of inference may introduce fallacies with respect to MTh. 1.1.2.6-7 unless closely watched.

Example:

An arithmetic theorem asserts that there is no largest integer, i.e.,

$$(x)(Ey)(x < y) \quad (1)$$

from which it may follow, by applying U.S.,

$$(Ey)(x < y) \quad (2)$$

from which in turn, by applying E.S., there follows

$$x < a \quad (3)$$

such that, by applying E.G.,

$$(Ex)(x < x) \quad (4)$$

which is of course false. The trouble started obviously in (2), where x is free. First of all, the variable should have been flagged and “(2)” should have been replaced by “ $x, (2)$ ” which explicitly shows that x is flagged in (2); likewise “ $x, (3)$ ” and “ $x, (4)$ ” instead of plain “(3)” and “(4)”, since they depend on (2).

To eliminate fallacies caused by MTh. 1.1.2.6-7, then, the following rule should be obeyed:

- (i) U.G. in deriving $(x)f(x)$ from $f(x)$ is valid iff x is not flagged in $f(x)$.

This rule alone is not enough, however, to safeguard an inference against fallacies. As is evident in the above example, there is another false step in (3), where an ambiguous term is carelessly introduced. Hence the second rule is:

- (ii) Any term introduced by E.S. must have subscripts to indicate all the free variables of the premises at issue.

In (3), for instance, " $x < a$ " should have been written as " $x < a_x$ " which yields " $(Ex)(x < a_x)$ " in (4) to be correct.

Certain ramifications may take place with respect to such rules of inference for a full and rigorous treatment, some of which may appear in a form of notes in the following problems, but it is certainly not the task of the present text to cover all of them.

Solved Problems

1. Find free variables in the following propositional functions:

- | | | |
|-----------------------------------|---|--|
| (i) $(x)f(x) \vee g(x)$ | (iv) $(x)\overline{A(x)B(y)}$ | (vii) $(Ex)F(x) \equiv G(x)$ |
| (ii) $(y)\overline{A(y)}$ | (v) $(x)[P(x) \rightarrow Q(x) \vee \overline{Q(x)}]$ | (viii) $(z)(Ey)[A(z)B(z)] \rightarrow C(x, y)$ |
| (iii) $(Ez)F(z) \rightarrow G(z)$ | (vi) $A(x) \rightarrow (x)B(x)$ | |

Solution:

- | | | |
|------------------------------|--|--------------------------------------|
| (i) x of $g(x)$ is free. | (iv) y of $\overline{B(y)}$ is free. | (vii) x of $G(x)$ is free. |
| (iii) z of $G(z)$ is free. | (vi) x of $A(x)$ is free. | (viii) x, y of $C(x, y)$ are free. |

2. Prove: (i) $\overline{(x)[P(x) \rightarrow Q(x)]} \equiv (Ex)[P(x)\overline{Q(x)}]$, (ii) $\overline{(Ex)[P(x)Q(x)]} \equiv (x)[P(x) \rightarrow \overline{Q(x)}]$.

PROOF:

- | | | |
|---|--|----------------------|
| (i) $\overline{(x)[P(x) \rightarrow Q(x)]}$ | $\equiv (Ex)[\overline{P(x) \rightarrow Q(x)}]$ | MTh. 1.1.2.5 |
| | $\equiv (Ex)[\overline{P(x)} \vee Q(x)]$ | §1.1.1, Prob. 9 |
| | $\equiv (Ex)[P(x)\overline{Q(x)}]$ | §1.1.1, Prob. 12, i |
| (ii) $\overline{(Ex)[P(x)Q(x)]}$ | $\equiv (x)[\overline{P(x)Q(x)}]$ | MTh. 1.1.2.5 |
| | $\equiv (x)[\overline{P(x)} \vee \overline{Q(x)}]$ | §1.1.1, Prob. 12, ii |
| | $\equiv (x)[P(x) \rightarrow \overline{Q(x)}]$ | §1.1.1, Prob. 9, i |

3. Symbolize the following inference, then justify each step of the inference: "All mammals are animals; some mammals are biped; therefore, some animals are biped."

PROOF:

Let M , A , and B represent the predicates of being mammal, animal, and biped respectively, and "Hyp" below, as everywhere else, will denote a hypothesis; then the symbolized inference runs as follows:

- | | |
|----------------------------------|--------------------------------|
| (1) $(x)[M(x) \rightarrow A(x)]$ | Hyp ₁ |
| (2) $(Ex)[M(x)B(x)]$ | Hyp ₂ |
| (3) $M(a)B(a)$ | E.S. twice in (2) |
| (4) $M(a) \rightarrow A(a)$ | U.S. in (1) |
| (5) $A(a)B(a)$ | (3), (4) and MTh. 1.1.1.11, 14 |
| (6) $(Ex)[A(x)B(x)]$ | E.G. in (5) |

4. Deduce, symbolically and justifiably, “Some periodic functions are continuous” from “All trigonometric functions are periodic functions” and “Some trigonometric functions are continuous”.

PROOF:

Let T , P , and C designate the attributes of trigonometric functions, periodic functions, and continuous functions respectively; then

(1)	$(x)[T(x) \rightarrow P(x)]$	Hyp ₁
(2)	$(Ex)[T(x)C(x)]$	Hyp ₂
(3)	$T(a)C(a)$	E.S. in (2)
(4)	$T(a) \rightarrow P(a)$	U.S. in (1)
(5)	$P(a)C(a)$	(3), (4), and MTh. 1.1.1.11, 14
(6)	$(Ex)[P(x)C(x)]$	E.G. in (5)

5. Given two premises: “All irrational numbers are real numbers” and “All real numbers are complex numbers”, draw a conclusion through a symbolic procedure of inference.

Solution:

If I , R , and C designate the predicates of being an irrational number, a real number, and a complex number, then

(1)	$(x)[I(x) \rightarrow R(x)]$	Hyp ₁
(2)	$(x)[R(x) \rightarrow C(x)]$	Hyp ₂
(3)	$I(a) \rightarrow R(a)$	U.S. in (1)
(4)	$R(a) \rightarrow C(a)$	U.S. in (2)
(5)	$I(a) \rightarrow C(a)$	(3), (4), and MTh. 1.1.1.13
(6)	$(x)[I(x) \rightarrow C(x)]$	U.G. in (5)

Note. U.G. in (5) is allowed only because the inference began with (x) and went through onto (5) without free x .

6. Symbolize, then justify, the following reasoning: “No rational being is willing to destroy the world; no maniac is unwilling to destroy the world; every sane person is a rational being; therefore, no sane person is a maniac.”

PROOF:

Let R , W , M , and S represent “rational being”, “willing to destroy the world”, “maniac”, “sane person”; then

(1)	$(x)[R(x) \rightarrow \overline{W(x)}]$	Hyp ₁
(2)	$(x)[M(x) \rightarrow W(x)]$	Hyp ₂
(3)	$(x)[S(x) \rightarrow R(x)]$	Hyp ₃
(4)	$R(a) \rightarrow \overline{W(a)}$	U.S. in (1)
(5)	$M(a) \rightarrow W(a)$	U.S. in (2)
(6)	$S(a) \rightarrow R(a)$	U.S. in (3)
(7)	$S(a) \rightarrow \overline{W(a)}$	(6), (4), and MTh. 1.1.1.13
(8)	$\overline{W(a)} \rightarrow \overline{M(a)}$	(5) and MTh. 1.1.1.12
(9)	$S(a) \rightarrow \overline{M(a)}$	(7), (8), and MTh. 1.1.1.13
(10)	$(x)[S(x) \rightarrow \overline{M(x)}]$	U.G. in (9)

7. Find fallacious steps in the following reasoning where O predicates an odd number:

(1)	$(Ex)\overline{O(x)}$	Hyp ₁
(2)	$\overline{O(a)}$	E.S. in (2)
(3)	$O(x)$	Hyp ₂
(4)	$\overline{O(a)}O(x)$	(2), (3), and MTh. 1.1.1.14
(5)	$(Ex)\overline{O(x)}O(x)$	E.G. in (4)

PROOF:

The concluded fallacy that there exists a number x such that it is odd and not odd was first introduced in (3), where x should have been flagged, then in (4), where x should have been flagged again, since it depends on (3); finally, (4) should never have been existentially generalized, since a and free x occur together in (4) (cf. MTh. 1.1.2.7, ii).

8. Justify, formally, the following reasoning: "All integers are rational numbers; therefore, all negative integers are negative rational numbers."

PROOF:

Let I , R , and N predicate integers, rational numbers, and negative respectively; then

(1)	$(x)[I(x) \rightarrow R(x)]$	Hyp ₁
(2)	$(Ey)[I(y)N(xy)]$	x , Hyp ₂
(3)	$I(a_x)N(xa_x)$	x , E.S. in (2)
(4)	$I(a_x) \rightarrow R(a_x)$	U.S. in (1)
(5)	$R(a_x)N(xa_x)$	x , (3), (4), and MTh. 1.1.1.11
(6)	$(Ey)[R(y)N(xy)]$	x , E.G. in (5)
(7)	$(Ey)[I(y)N(xy)] \rightarrow (Ey)[R(y)N(xy)]$	(2)-(6), and MTh. 1.1.1.13
(8)	$(x)(Ey)[I(y)N(xy)] \rightarrow (Ey)[R(y)N(xy)]$	U.G. in (7)

9. Symbolize, then justify, the following inference: "None of the primes are integrally divisible by an even integer greater than 2; any of the primes is integrally divisible by the unity; there exist some primes; therefore, the unity is not integrally divisible by an even integer greater than 2."

PROOF:

Let P , D , I , and U denote "primes", "is integrally divisible by", "even integers greater than 2", and "the unity" respectively; then

(1)	$(x)[P(x) \rightarrow (y)[I(y) \rightarrow \overline{D(xy)}]]$	Hyp ₁
(2)	$(x)[P(x) \rightarrow (Ey)[U(y)D(xy)]]$	Hyp ₂
(3)	$(Ex)[P(x)]$	Hyp ₃
(4)	$P(a)$	E.S. in (3)
(5)	$P(a) \rightarrow (y)[I(y) \rightarrow \overline{D(ay)}]$	U.S. in (1)
(6)	$P(a) \rightarrow (Ey)[U(y)D(ay)]$	U.S. in (2)
(7)	$(y)[I(y) \rightarrow \overline{D(ay)}]$	(4), (5), and MTh. 1.1.1.11
(8)	$I(b) \rightarrow \overline{D(ab)}$	U.S. in (7)
(9)	$(Ey)[U(y)D(ay)]$	(4), (6), and MTh. 1.1.1.11
(10)	$U(b)D(ab)$	E.S. in (9)
(11)	$\overline{I(b)}$	(8), (10), and MTh. 1.1.1.12
(12)	$U(b)\overline{I(b)}$	(10), (11), and MTh. 1.1.1.14
(13)	$(Ex)[U(x)\overline{I(x)}]$	E.G. in (12)

Note. $U(x)$ being by definition one and only one number 1 among integers, the conclusion is not actually for "some x " representing more than one number; in other contexts, however, it may be literally for "some x ".

Chapter 1.2

*Mathematical Proofs

Df. 1.2.1 $P_1, P_2, \dots, P_n \vdash P$, which reads “ P_1, P_2, \dots, P_n yield P ”, means that a proposition P is finally derived from a *sequence* of other propositions P_1, P_2, \dots, P_n .

Example:

Prob. 15-18 of §1.1.1 and Prob. 2-9 of §1.1.2 are treated by the sequences of several propositions, sometimes as many as seventeen, to arrive at the final proposition to be justified. In Prob. 9, for instance, P_n is the twelfth proposition and P the thirteenth.

The symbol “ \vdash ” is called a turnstile, denoting an assertion.

Df. 1.2.2 $\vdash P$, which reads “yields P ”, means that P is derived directly from axioms, and from them alone, by MTh. 1.1.1.9-11.

Example:

Prob. 15, ii, iii, iv of §1.1.1.

Df. 1.2.3 A *demonstration*, which may be symbolically represented by $D_1, D_2, \dots, D_m \vdash D$ where D is a proposition to appear at the end as a consequence of the propositions D_1, D_2, \dots, D_m , is a sequence of propositions P_1, P_2, \dots, P_n such that $P_i, i=1, 2, \dots, n$, is either an axiom or any of the D 's or any of the $P_j, j < i$, or whatever is derived from the two preceding P 's by MTh. 1.1.1.9-11.

Example:

Prob. 9 of §1.1.2 has $\text{Hyp}_1, \text{Hyp}_2, \text{Hyp}_3$ as D_1, D_2, D_3 and $(Ex)(U(x)\overline{I(x)})$ as D ; all other steps represent P_i or the result of the application of MTh. 1.1.1.9-11. Note that the seventh step, for instance, may be in need of a long demonstration for itself to justify the logical inference involved in the step.

As is obvious even in a single example, demonstrations may not, and sometimes technically cannot, always be carried out in full detail, since they are generally of staggering length except for exceptionally simple problems as Prob. 15-18 of §1.1.1. It is not only impractical and unnecessarily tedious, but not even desirable to write out every detail of an entirety of logical reasoning for each problem, let alone an analysis or justification of each step in the logical reasoning.

In practice, therefore, demonstrations must naturally suffer from certain, often drastic, abridgments, the amount of which depends on their prospective readers. There is no harm, of course, in such abridgments as long as it is understood that the demonstrations, on demand, can fill in all missing steps. For the sake of convenience and practicality, therefore, the following definition is accepted, if only tacitly, by all working mathematicians.

Df. 1.2.4 A *mathematical proof* is a set of representative clues, intelligible to whom it is intended, which point to the existence of a demonstration.

It is because of this reason that proofs in advanced research papers usually omit so much of details that they can be considered intelligible by but few experts in the field. When challenged, however, the writers of such papers may go all lengths to fill in omitted steps or give detailed demonstrations for certain unintelligible parts.

However abridged a mathematical proof may be, a proof as a model of precision must always meet the specification inherent in the core of demonstrations, viz. the logical inference from the assumed (hypotheses) to the justified (conclusions), since mathematics itself, as knowledge, must always proceed from what is given to what is to be verified or justified, or more broadly, from the known to the unknown. Such a procedure is of necessity presented in the form of implications (or what is the same, conditionals).

Df. 1.2.5 If a proposition A implies a proposition B , i.e. $A \rightarrow B$, then B is said to be a *necessary condition* for A .

Stated otherwise: “a necessary condition that A be true is that B be true” means “ A implies B ” or “If A , then B ” or “ B only if A ”.

Example:

A necessary condition that an integer be integrally divisible by 4 is that it be integrally divisible by 2; a necessary condition that a quadrilateral be a rectangle is that it be a parallelogram. Note that, as in these examples, necessary conditions connote *minimal* conditions.

Df. 1.2.6 If a proposition A implies a proposition B , then A is said to be a *sufficient condition* for B .

Stated otherwise: “a sufficient condition that B be true is that A be true” means “ A implies B ” or “If A , then B ” or “ B only if A ”.

Example:

A sufficient condition that an integer be integrally divisible by 4 is that it be integrally divisible by 8; a sufficient condition that a quadrilateral be a rectangle is that it be a square. *Maximal* conditions thus connote sufficient conditions.

In an abstract context the line of demarcation between necessary and sufficient conditions may not look sufficiently distinct, since A is a sufficient condition for B and B is a necessary condition for A whenever A implies B , but the line becomes quite clear in a concrete context.

Example:

“If n is integrally divisible by 4, then it is integrally divisible by 2” is quite distinguishable from “If n is integrally divisible by 2, then it is integrally divisible by 4”; the latter is obviously false while the former is true. In the former the if-clause (A) is indeed a sufficient condition for the then-clause (B), and B is a necessary condition for A .

Df. 1.2.7 If a proposition A implies a proposition B which in turn implies A , then A (or B) is said to be a *necessary and sufficient condition* for B (or A), or A and B are said to be *logically equivalent*.

Stated otherwise: “a necessary and sufficient condition that A be true is that B be true” means “ A implies B and B implies A ” or “ A implies B , and conversely” or “ A and B are logically equivalent” or “ A iff B ” or “ B iff A ”.

This is the only case where a necessary condition is also a sufficient condition, and vice versa. E.g., a necessary and sufficient condition that an integer n be integrally divisible by 4 is that n be a multiple of 4; a necessary and sufficient condition that a quadrilateral be a rectangle is that it be a parallelogram with one angle a right angle.

A necessary and sufficient condition in a definite context may be stated in several different ways as long as the results are all logically equivalent themselves; e.g., a necessary and sufficient condition that an integer be integrally divisible by 9 is that it be a multiple of 9 or that the sum of its digits be a multiple of 9.

It must be emphasized here again that the necessary or sufficient condition with respect to mathematical proofs is certainly not physical, but purely logical, in the well-defined sense that the “if-then” connective is not necessarily of a causal relation, and that “ $A \rightarrow B$ ” is logically equivalent to “ $\bar{A} \vee B$ ” or “ $\overline{A\bar{B}}$ ”.

In general, an implication with respect to two propositions A and B is studied in the frame of proofs by the following four cases:

- (i) $A \rightarrow B$ (ii) $B \rightarrow A$ (iii) $\bar{A} \rightarrow \bar{B}$ (iv) $\bar{B} \rightarrow \bar{A}$

Df. 1.2.8 Given an implication $A \rightarrow B$, its *converse* is $B \rightarrow A$, its *opposite* (or *inverse*) $\bar{A} \rightarrow \bar{B}$, and its *contrapositive* (or *opposite converse*) $\bar{B} \rightarrow \bar{A}$.

Example:

If A and B represent two propositions “ T_1 and T_2 are two similar triangles” and “the corresponding angles of the two triangles T_1 and T_2 are equal” respectively, then

- (i) $A \rightarrow B$: “if T_1 and T_2 are two similar triangles, then the corresponding angles of the two triangles T_1 and T_2 are equal.”
- (ii) $B \rightarrow A$: “if the corresponding angles of the two triangles T_1 and T_2 are equal, then T_1 and T_2 are two similar triangles.”
- (iii) $\bar{A} \rightarrow \bar{B}$: “if T_1 and T_2 are not two similar triangles, then the corresponding angles of the two triangles T_1 and T_2 are not equal.”
- (iv) $\bar{B} \rightarrow \bar{A}$: “if the corresponding angles of the two triangles T_1 and T_2 are not equal, then T_1 and T_2 are not two similar triangles.”

It just happens in this special case that everyone of the four alternatives is true, simply because A and B are logically equivalent; it will still be the same if A and B read “two lines are parallel” and “the two lines do not intersect” respectively, within the frame of reference definitely defined as Euclidean space. These cases, however, are exceptional, and as has already been exemplified by “ n is an integer integrally divisible by 4” and “ n is integrally divisible by 2”, it is usually the case that the converse of a proposition does not hold even if the proposition holds. Hence the following distinction:

MTh. 1.2.9 If “ $A \rightarrow B$ ” is a theorem, so is always “ $\bar{B} \rightarrow \bar{A}$ ”, but not always “ $B \rightarrow A$ ” and “ $\bar{A} \rightarrow \bar{B}$ ”; on the other hand, if “ $B \rightarrow A$ ” is a theorem, so is always “ $\bar{A} \rightarrow \bar{B}$ ”, but not always “ $A \rightarrow B$ ” and “ $\bar{B} \rightarrow \bar{A}$ ”.

Example:

A theorem in Euclidean geometry: “If two lines are parallel, then the lines do not intersect” may be legitimately established by proving its opposite converse: “If two lines intersect, the lines are not parallel.” Likewise, using the example of MTh. 1.1.1.11: “if an infinite series converges, then the general term of the series approaches zero” is logically equivalent to: “if the general term of an infinite series does not approach zero, then the given series does not converge”. In either form the theorem may be proved to be true and then applied to other problems, but it does not logically follow from the theorem that “if the general term of an infinite series approaches zero, then the given series converges” (which is generally false) or “if an infinite series does not converge, then the general term of the series does not approach zero” (which, again, is generally false).

Note the similarity between this metatheorem and Prob. 13 of §1.1.1; in the same spirit, it can be readily verified by truth-tables that “ $p \rightarrow q$ ” is not logically equivalent to “ $q \rightarrow p$ ” or “ $\bar{p} \rightarrow \bar{q}$ ”.

Other modes of indirect proofs are also available as follows.

MTh. 1.2.10 “ $A \rightarrow B$ ” is a theorem if a contradiction “ CC ” can be derived from “ $A \rightarrow \bar{B}$ ”.

Stated otherwise: if the negation of the desired conclusion (B) is introduced as a new hypothesis (\bar{B}), and if the use of this new hypothesis (\bar{B}) together with the original hypothesis (or hypotheses, represented by A in either case) brings forth a contradiction (CC), then it must be the case that $A \rightarrow B$.

This metatheorem is merely a symbolized form of the so-called *indirect proof* or *reductio ad absurdum proof*, whose examples are abundant in elementary geometry (with which the reader is quite familiar).

Note the resemblance in the form of reasoning between this metatheorem and Prob. 6, iii of §1.1.1; note, also, the way this metatheorem was already applied to Prob. 18 of §1.1.1, in particular to the steps (10)-(12). In this sense the metatheorem may be considered also patterned after the familiar tautology: $p \rightarrow q \equiv \bar{p}\bar{q}$; the reasoning in the present context, then, may be symbolized as follows: $p\bar{q} \rightarrow r\bar{r} \rightarrow \bar{p}\bar{q} \rightarrow (p \rightarrow q)$.

MTh. 1.2.11 If $A_1 \vee A_2 \vee \dots \vee A_n$, and if $\bar{A}_1, \bar{A}_2, \dots$, and \bar{A}_{n-1} , then A_n is a theorem.

Example:

$x = y$ if it can be proved that neither $x < y$ nor $x > y$ where the frame of reference is the trichotomy: $x < y$ or $x = y$ or $x > y$. As this example verifies, it must be assumed that the alternatives A_1, A_2, \dots, A_n exhaust the entire case.

Symbolized in terms of propositional calculus, this metatheorem has the following form (which can be readily verified by truth-tables):

$$(p \vee q \vee r)(\bar{p}\bar{q}) \rightarrow r \quad \text{or in general} \quad (p_1 \vee p_2 \vee \dots \vee p_n)(\bar{p}_1 \bar{p}_2 \dots \bar{p}_{n-1}) \rightarrow p_n$$

It is self-explanatory that MTh. 1.2.9-10 may be freely employed to deduce $\bar{p}_1, \bar{p}_2, \dots, \bar{p}_{n-1}$ individually.

It must be emphasized as a general remark that a single counter-example is quite sufficient to *disprove* a case.

(As for the modes of fallacious reasonings, such as *petitio principii*, *non sequitur*, *post hoc ergo propter hoc*, etc., they are found in any text-book, old and new, on Logic.)

Supplementary Problems

Part 1

TAUTOLOGIES

- 1.1. Prove, by truth-tables, the following tautologies: (i) $pq \rightarrow p$, (ii) $pq \rightarrow p \vee q$.
- 1.2. Prove that the following twofold distributions under disjunction and conjunction are tautologies:
 - (i) $(p \vee q)(r \vee s) \equiv pr \vee qr \vee ps \vee rs$
 - (ii) $pq \vee rs \equiv (p \vee r)(q \vee r)(p \vee s)(q \vee s)$
- 1.3. Negation of equivalent terms is equivalent to the original equivalence; i.e. $(p \equiv q) \equiv (\bar{p} \equiv \bar{q})$ is a tautology.
- 1.4. Prove the redundancy of a negation: $p \vee \bar{p}q \equiv p \vee q$.

- 1.5. Complete the replacement of the five connectives by the stroke and the dagger, defined by Prob. 2-3 of §1.1.1; i.e., express the secondary connectives \rightarrow and \leftrightarrow also in terms of $|$ and \downarrow .
- 1.6. Express $p | q$ in terms of \downarrow and vice versa, then justify the expression by truth-tables.
- 1.7. Prove that a proposition which implies its own negation is a contradiction.
- 1.8. Find the exact relation between the following pair of propositions: $\bar{a} \vee \bar{b}\bar{c}$ and $a \rightarrow (a(\bar{b} \vee c))$.
- 1.9. Prove the following tautologies without using truth-tables:
- (i) $\bar{a}((\bar{b} \vee c)(d\bar{e})) \equiv a \vee \bar{b} \vee c \vee \bar{d} \vee e$ (ii) $(a \vee b \vee c)(\bar{a} \vee \bar{b} \vee c)(a \vee b \vee \bar{c}) \equiv (\bar{a}\bar{b}c) \vee (\bar{a}b\bar{c}) \vee (ab\bar{c})$
- 1.10. Verify the following tautologies, first by truth-tables, then by MTh. 1.1.1.9-10:
- (i) $(a \equiv b) \rightarrow (a \vee c \equiv b \vee c)$ (iii) $(a \equiv b) \rightarrow (a \rightarrow c \equiv b \rightarrow c)$
(ii) $(a \equiv b) \rightarrow (ac \equiv bc)$ (iv) $(a \equiv b) \rightarrow ((a \equiv c) \equiv (b \equiv c))$
- 1.11. Prove, first by truth-tables, then without truth-tables: $((a \rightarrow \bar{b}) \rightarrow (a \equiv \bar{b})) \equiv a \vee b$.
- 1.12. Test, by truth-tables, the validity or fallacy of the following propositions:
- (i) $(a \rightarrow b)(\bar{b} \rightarrow \bar{c}) \rightarrow (c \rightarrow a)$ (ii) $(a \rightarrow b)(\bar{c} \rightarrow \bar{b}) \rightarrow (\bar{c} \rightarrow \bar{a})$ (iii) $(a \vee b)(\bar{a} \vee \bar{c})(\bar{b} \rightarrow \bar{c}) \rightarrow \bar{a}$
- 1.13. Deduce, by MTh. 1.1.1.9-11 and Prob. 15-17, the tautology: $(p \rightarrow q) \rightarrow (\bar{q}r \rightarrow \bar{p}r)$.
- 1.14. Is \bar{a} deducible from three hypotheses: $a \rightarrow b$, $\bar{b} \vee c$, $\bar{a}\bar{c}$? If so, justify the logical inference (without any use of truth-tables).
- 1.15. Deduce, using only metatheorems, \bar{a} from three hypotheses: $ab \rightarrow cd$, b , \bar{d} .

QUANTIFICATIONS

- 1.16. Prove the following quantified tautologies:

$$(i) (Ex)(P(x)Q(x)) \rightarrow (Ex)P(x) \cdot (Ex)Q(x) \qquad (ii) (x)(P(x)Q(x)) \leftrightarrow (x)P(x) \cdot (x)Q(x)$$

- 1.17. Discuss the fallacy involved in the inference: $(Ex)P(x) \cdot (Ex)Q(x) \rightarrow (Ex)(P(x)Q(x))$.

- 1.18. Prove: $(x)P(x) \rightarrow (Ex)P(x)$.

- 1.19. Prove the following quantified tautologies:

$$\begin{aligned} (i) & (x)P(x) \vee (x)Q(x) \rightarrow (x)(P(x) \vee Q(x)) \\ (ii) & (x)(P(x) \vee Q(x)) \rightarrow (x)P(x) \vee (Ex)P(x) \\ (iii) & (Ex)(P(x) \vee Q(x)) \leftrightarrow (Ex)P(x) \vee (Ex)Q(x) \\ (iv) & (x)(P(x) \rightarrow Q(x)) \rightarrow ((x)P(x) \rightarrow (x)Q(x)) \\ (v) & (x)(P(x) \rightarrow Q(x)) \rightarrow ((Ex)P(x) \rightarrow (Ex)Q(x)) \\ (vi) & (Ex)(P(x) \rightarrow Q(x)) \leftrightarrow (x)P(x) \rightarrow (Ex)Q(x) \end{aligned}$$

- 1.20. Given the well-established theorem that the square of an even integer is again an even integer, symbolize the proof that an integer is odd if its square root is also an odd integer.

Sets in General

Df. 2.1.1 A *set* is a well-defined collection of distinct elements.

This definition, where the “set” is defined in terms of its synonym “collection”, is obviously nominal. In this sense, as is well known, the set cannot be properly defined, although it may be replaced by any of its synonyms such as collection, class, aggregate or even family and can be readily exemplified by any of collective nouns such as army, assembly, flock, herd, jury, etc.

The term “well-defined” in Df. 2.1.1, however, specifies that it can be determined at least whether or not certain *elements* belong to the set in question (where the elements themselves remain undefined), and the term “distinct” specifies that, given two elements, their identity or difference can be discerned. Such a discernment is considered always possible in logic and mathematics on the strength of the most fundamental metatheorem, which runs as follows:

MTh. 2.1.1a (Principle of Identity). Whatever is is *identical* with itself (cf. §1.1.1, Prob. 5 note).

Example: A set S is identical with S itself.

Df. 2.1.1b The membership of a set is denoted by “ \in ” and the non-membership by “ \notin ”.

Example:

$x \in X$ designates that x is a member of a set X and reads “ x is an element of X ” or “ x belongs to X ”, while $x \notin X$ reads “ x is not an element of X ” or “ x does not belong to X ”. It is customary to use small letters for elements and capital letters for sets. If N denotes the set of all natural numbers, then $x \in N$ specifies that x is a natural number, and $y \notin N$ designates that y may be a negative integer or an irrational number or anything but a natural number.

Since a set is uniquely determined by its elements, the elements of the set, enclosed in braces, may be explicitly listed as a notation for the set itself; e.g. $A = \{a, b, c\}$ for a set A whose elements are a, b, c and nothing else. If B is a set which consists of a, b, c and possibly more, then notationally, $B = \{a, b, c, \dots\}$. On the other hand, a set which consists of a single element is called a *unit set*, and the set whose only element is x is sometimes called *singleton x* , denoted by $\{x\}$.

Df. 2.1.2 If each element of a set X is also an element of a set Y , then X is called a *subset* of Y , denoted by $X \subseteq Y$ which reads “ X is contained in Y ” (or what is the same: $Y \supseteq X$, which reads “ Y contains X ”).

Example:

$N \subseteq I$, if N is the set of all natural numbers and I the set of all integers; in this context, N is a (non-empty) subset of I .

Th. 2.1.3 If $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C$. (Cf. Prob. 1.)

Example: If $N \subseteq I$ and $I \subseteq R$ where R is the set of all rational numbers, then $N \subseteq R$.

MTh. 2.1.4 (Axiom of Extension). Two sets X and Y are *equal*, denoted by $X = Y$, iff they have the same elements.

Example:

If $A = \{a, b, c\}$ and $B = \{a, c, b\}$, then $A = B$. It must be noted that the *equal* sets may not be *identical* (cf. MTh. 2.1.1a), as is actually the case here.

Th. 2.1.5 If $X \subseteq Y$ and $Y \subseteq X$, then $X = Y$. (Cf. Prob. 2.)

Df. 2.1.6 A set of certain elements in its entirety is called the *universal set* (or *universe*), denoted by U , which is considered a subset of U itself.

Example:

Given $A = \{a, b, c\}$ alone with respect to itself, then A is U ; it is also a subset of itself, since it satisfies Df. 2.1.2.

Df. 2.1.7 A set which does not contain any element at all is *empty* (or *vacuous*) and called the *null set*, denoted by \emptyset ; it is considered a subset of every set.

Example:

Taking away a, b, c from $A = \{a, b, c\}$, A becomes empty, i.e. \emptyset ; \emptyset is a subset of, say, $C = \{c, b, a\}$, since each member (nothing!) of \emptyset belongs to C .

Note. Df. 1.1.2.3 gives Df. 2.1.7 a more formal expression, viz. $(x)(x \neq x)$ or what is the same: $\{x: x \neq x\}$ or $\{x | x \neq x\}$.

Df. 2.1.8 If a set X is a subset of a set Y and at least one element of Y is not an element of X , then X is called a *proper subset* of Y , denoted by $X \subset Y$; the null set is considered a proper subset of every set except itself.

Example:

The proper subsets of $\{a, b, c\}$ are: $\{a, b\}$, $\{a, c\}$, $\{b, c\}$, $\{a\}$, $\{b\}$, $\{c\}$, and \emptyset . Including itself $\{a, b, c\}$, which is a subset of itself, although definitely not a proper subset of itself, the number of the subsets of the set is $2^3 = 8$, which gives the following generalization.

Th. 2.1.9 A set S of n elements has 2^n subsets. (Cf. Prob. 7.)

Df. 2.1.10 Given two sets X and Y , there exists a *one-to-one* (or *1-1*) *correspondence* between X and Y iff $x \in X$ has one and only one correspondent $y \in Y$, denoted by $x \leftrightarrow y$ or $x \in X \leftrightarrow y \in Y$.

Example:

If X denotes the set of all positive integers and Y the set of all negative integers, then there exists a 1-1 correspondence between X and Y , since $1 \leftrightarrow -1$, $2 \leftrightarrow -2$, ..., $n \leftrightarrow -n$, ...

Df. 2.1.11 If the elements of two sets X and Y can be placed in one-to-one correspondence, they are said to be (*cardinally*) *equivalent*, denoted by $X \leftrightarrow Y$; they are also said to have the same *cardinal number*, denoted by $o(X) = o(Y)$.

Example:

There exists a 1-1 correspondence between the set X of all positive integers and the set Y of all positive odd integers; hence $o(X) = o(Y)$, viz.,

X:	1	2	3	4	5	.	.	.	100	.	.	n	.	.
	\updownarrow	\updownarrow	\updownarrow	\updownarrow	\updownarrow				\updownarrow			\updownarrow		
Y:	1	3	5	7	9	.	.	.	199	.	.	$2n-1$.	.

Df. 2.1.12 The relation of equivalence satisfies the following three laws, called the *equivalence relations*:

Reflexive:	$X \leftrightarrow X$
Symmetric:	$X \leftrightarrow Y$ implies $Y \leftrightarrow X$
Transitive:	$X \leftrightarrow Y$ and $Y \leftrightarrow Z$ imply $X \leftrightarrow Z$

Example:

The relation represented by the logical equivalence (cf. Df. 1.1.1.6) is an equivalence relation, since

- (i) $p \equiv p$ (ii) $p \equiv q$ implies $q \equiv p$ (iii) $p \equiv q$ and $q \equiv r$ imply $p \equiv r$

Df. 2.1.13 A set has cardinal number n , viz. $o(S) = n$, iff there exists a 1-1 correspondence between the elements of S and the integers $1, 2, 3, \dots, n$; such a set is *finite*.

Example:

$C = \{c, b, a\}$ is of cardinal number 3, i.e. $o(C) = 3$, which is obviously a finite number.

Df. 2.1.14 A set S is *countable* (or *denumerable*) and has cardinal number d , viz. $o(S) = d$ (or the so-called “aleph null”), iff there exists a 1-1 correspondence between the elements of S and all positive integers; such a set is *infinite*.

E.g. cf. Th. 2.1.15 below.

Th. 2.1.15 The set of all rational numbers is a countable set, but the set of all real numbers is not. (Cf. Problems 12-14)

Df. 2.1.16 A set S which is equivalent to the set of all real numbers is said to have the cardinal number c of the continuum, viz. $o(S) = c$ (or the so-called “aleph-one”).

Example:

The set of all points in a closed interval $[0, 1]$ (cf. Problems 9, 10, 14).

Solved Problems

1. If $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C$.

PROOF:

Take any element x which belongs to A , viz. $x \in A$; then $x \in B$ since $A \subseteq B$. Then also $x \in C$ since $B \subseteq C$. Hence every element of A is also an element of C , i.e. $A \subseteq C$.

2. If $A \subseteq B$ and $B \subseteq A$, then $A = B$.

PROOF:

Let $x \in A$; then $x \in B$ since $A \subseteq B$. Conversely, if $x \in B$, then $x \in A$ since $B \subseteq A$. Hence, having the same elements, $A = B$.

3. If $A \subset B$ and $B \subseteq C$, then $A \subset C$.

PROOF:

Since $A \subset B$ and $B \subseteq C$, at most $A \subseteq C$. But if $A = C$, then $B \subseteq A$ ($\because B \subseteq C$), which is contradictory to $A \subset B$. Hence $A \subset C$.

4. The set R of all rational numbers is a proper subset of the set R^* of all real numbers.

PROOF:

Since every rational number is also a real number, $R \subseteq R^*$. There exists, however, at least one element of R^* , say $\sqrt{2}$, which does not belong to R . Hence $R \neq R^*$, and from $R \subseteq R^*$ and $R \neq R^*$ it follows that $R \subset R^*$.

(Note that the existence of any irrational number is enough for the proof; instead of $\sqrt{2}$, other well-known irrational numbers such as π , e , etc. may be used, although their irrationality is not so easy to prove as that of $\sqrt{2}$, which runs as follows: If $\sqrt{2}$ is a rational number, then $\sqrt{2} = p/q$ where p and q are positive integers without any common divisor but 1. Squaring both sides of the equation, $2q^2 = p^2$, meaning that p must be an even number. Hence let $p = 2p'$; then $2q^2 = 4p'^2$, i.e. $q^2 = 2p'^2$, meaning that q is also an even integer and that p and q do have a common divisor other than 1, contrary to the initial assumption. Hence $\sqrt{2}$ is not a rational number.)

5. In Mathematical Logic the statement "if p , then q " (cf. Df. 1.1.1.5) is true whenever either p and q are both true or p is false and q may be true or false. Using this logic, prove that the null set is unique.

PROOF:

Let \emptyset_1 and \emptyset_2 be two null sets, which must be proved to be equal, i.e. $\emptyset_1 = \emptyset_2$ or what is the same, $\emptyset_1 \subset \emptyset_2$ and $\emptyset_2 \subset \emptyset_1$. The former is proved when the statement "if $x \in \emptyset_1$, then $x \in \emptyset_2$ " is proved to be true. Since \emptyset_1 is vacuous, $x \in \emptyset_1$ is false, and the statement as a whole is always true, i.e. $\emptyset_1 \subset \emptyset_2$.

Likewise the latter is proved if the proposition "if $x \in \emptyset_2$, then $x \in \emptyset_1$ " is proved to be true. Since \emptyset_2 is vacuous by hypothesis, $x \in \emptyset_2$ is false, and the proposition as a whole is true, i.e. $\emptyset_2 \subset \emptyset_1$.

Hence, putting two conclusions together and by Th. 2.1.5 and Df. 2.1.8, $\emptyset_1 = \emptyset_2$, and the null set is unique.

6. Both Mathematical Logic and traditional Aristotelean Logic define the same contrapositive rule (cf. MTh. 1.2.1.12): "If p , then q " is equivalent to "if not q , then not p ". Prove by this rule that the null set is unique.

PROOF:

Let \emptyset_1 and \emptyset_2 be two null sets as above; then, since the statement "if $x \notin \emptyset_2$, then $x \notin \emptyset_1$ " is always true according to the definition of the null set, the contrapositive rule proves that "if $x \in \emptyset_1$, then $x \in \emptyset_2$ ", i.e. $\emptyset_1 \subseteq \emptyset_2$.

Likewise, since the proposition "if $x \notin \emptyset_1$, then $x \notin \emptyset_2$ " is true by definition, it is immediately deduced through the contrapositive rule that "if $x \in \emptyset_2$, then $x \in \emptyset_1$ ", i.e. $\emptyset_2 \subseteq \emptyset_1$.

Hence, taking both conclusions together and by Th. 2.1.5, $\emptyset_1 = \emptyset_2$.

7. A set S of n elements has 2^n subsets.

PROOF:

In general, the number of the subsets whose elements are m out of n is the number of combinations of n elements taken m at a time, that is,

$${}_nC_m = n! / (m!(n-m)!)$$

Hence the sum of all subsets, including the universe (${}_nC_n$) and the null set (${}_nC_0$), is

$$\begin{aligned} \sum_{m=0}^n {}nC_m &= {}nC_0 + {}nC_1 + \cdots + {}nC_n \\ &= (1+1)^n = 2^n \end{aligned}$$

Note. This proof presumes the binomial theorem (with which the reader is quite familiar); there appears in any textbook of the subject a theorem deducible from the binomial theorem that

$$2^n = (1+1)^n = {}nC_0 + {}nC_1 + \cdots + {}nC_r + \cdots + {}nC_n$$

8. The set I of all integers is equivalent to the set N of all natural numbers.

PROOF:

A 1-1 correspondence between I and N can be made as follows:

I :	0	1	-1	2	-2	.	.	.	m	$-m$.	.
	\updownarrow	\updownarrow	\updownarrow	\updownarrow	\updownarrow	.	.	.	\updownarrow	\updownarrow	.	.
N :	1	2	3	4	5	.	.	.	$2m$	$2 m + 1$.	.

9. Find a 1-1 correspondence in each pair of the following intervals: (i) two closed intervals of $[0, 1]$ and $[0, 2]$; (ii) $[0, 2]$ and a straight line of infinite length.

Solution:

Draw the projecting rays for (i) from a point A , and for (ii) from two points A and B , as in Fig. 2.1a and 2.1b. Then, as is rather self-explanatory in the figures themselves, the two points a and b in the interval $[0, 1] = \overline{BC}$ in Fig. 2.1a are the correspondents of a' and b' in the interval $[0, 2] = \overline{DE}$, i.e. $a \leftrightarrow a'$, $b \leftrightarrow b'$, and in general, for any two points $x \in \overline{BC}$ and $x' \in \overline{DE}$: $x \leftrightarrow x'$.

Likewise, in Fig. 2.1b, $a \leftrightarrow a'$, $b \leftrightarrow b'$, $c \leftrightarrow c'$, $d \leftrightarrow d'$, and in general, for any two points $x \in \overline{CD}$ and $x' \in \overline{EF}$, $x \leftrightarrow x'$, proving the desired 1-1 correspondence.

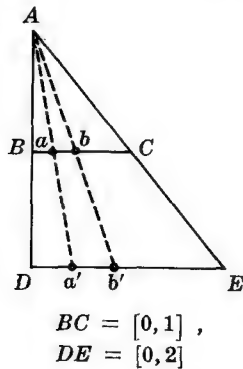


Fig. 2.1a

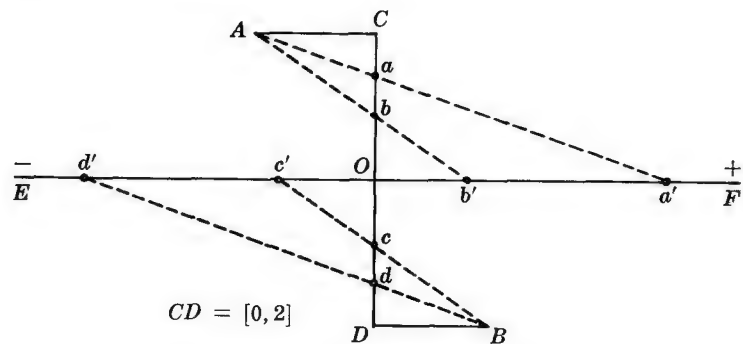


Fig. 2.1b

10. Find a 1-1 correspondence between a closed interval $[0, 1]$ (i.e. $0 \leq x \leq 1$) and a half-closed interval $[0, 1)$ (i.e. $0 \leq x' < 1$).

Solution:

As in Fig. 2.1c, let $x \leftrightarrow x'$ if $x \neq 1/2^n$, $n = 1, 2, \dots$; if $x = 1/2^n$, then let $x \leftrightarrow 1/2^{n+1} = x'$.

E.g., if $x = 1$, i.e. $x = 1/2^0$, then it is made 1-1 correspondent to $x' = 1/2^{0+1} = 1/2$, and $x = 1/2 \leftrightarrow x' = 1/4$, $x = 1/4 \leftrightarrow x' = 1/8$, etc., while the fraction of any other type, e.g. $x = 1/3$, is made directly 1-1 correspondent to itself, i.e. $x' = 1/3$.

Then, since any point in the interval $0 \leq x \leq 1$ is either the fraction of the type $1/2^n$ or the proper fraction of some other type, expressed by $x \neq 1/2^n$, the two modes of 1-1 correspondence exhaust all possible correspondences between x and x' , completing the proof.

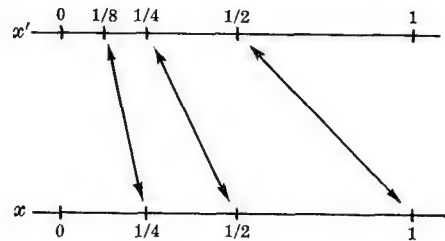


Fig. 2.1c

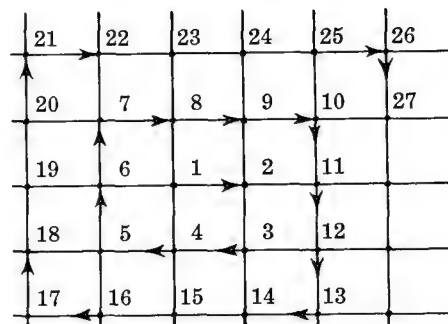
11. There exists a 1-1 correspondence between the set of all natural numbers and the set of all lattice points (i.e. the points whose coordinates are integers) in the plane.

PROOF:

It is self-explanatory in Fig. 2.1d below, since each lattice point is now made 1-1 correspondent to a natural number.

E.g., the lattice points $P_1, P_2, P_3, P_4, P_5, \dots$ may be now replaced simply by the natural numbers 1, 2, 3, 4, 5, \dots , since there obviously exist the 1-1 correspondences: $P_1 \leftrightarrow 1, P_2 \leftrightarrow 2$, etc. Furthermore, all the lattice points will certainly be exhausted by this procedure if the counting begins with 1 and moves in the direction pointed by arrows, i.e. $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow \dots$, completing the proof.

Fig. 2.1d



12. The set of all rational numbers is countable.

PROOF:

Cf. Problem 11 above, and note that the set of all lattice points can be put into 1-1 correspondence with the set of all rational numbers represented by the pairs of coordinates in the form of quotients (cf. Prob. 4), even when duplicates are omitted. Hence the latter also is countable.

Second proofs. The mode of 1-1 correspondence above is by no means unique; others, e.g. Fig. 2.1e, 2.1f, 2.1g, are also available.

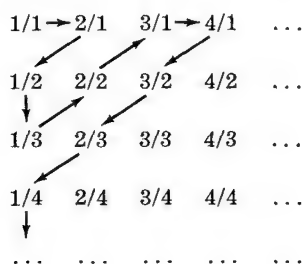


Fig. 2.1e

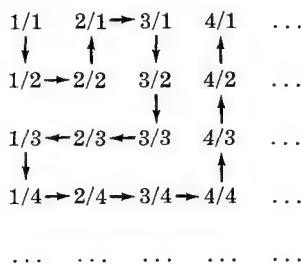


Fig. 2.1f

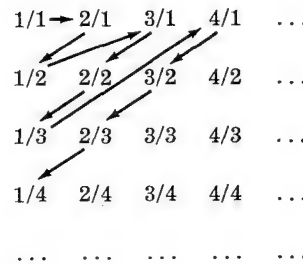


Fig. 2.1g

These proofs establish the 1-1 correspondence between the set of all natural numbers and the set of all *positive* rational numbers, but this can be readily extended to the set of all rational numbers (cf. Prob. 8).

13. A real algebraic number (cf. Df. 5.3.2.2) is a real root satisfying a polynomial equation with integral coefficients

$$a_0 x^n + a_1 x^{n-1} + \dots + a_{n-1} x + a_n = 0, \quad a_0 > 0, \quad n \geq 1;$$

and the set A of all real algebraic numbers, containing such irrational numbers as $\sqrt{2}$ and $\sqrt{3}$ (the roots of $x^2 - 2 = 0$, $x^3 - 3 = 0$ respectively), is larger than the set R of all rational numbers (i.e. $R \subset A$). Prove that A is nevertheless countable.

PROOF:

Consider the *height* h of the equation, defined by

$$h = n + a_0 + |a_1| + \dots + |a_{n-1}| + |a_n|$$

Since both n and a_0 are at least 1, $h \geq 2$. The equations of height 2, 3, 4 are respectively

$$x = 0; \quad 2x = 0, \quad x \pm 1 = 0, \quad x^2 = 0; \quad 3x = 0, \quad x \pm 2 = 0, \quad 2x \pm 1 = 0, \quad 2x^2 = 0, \quad x^2 \pm 1 = 0, \quad x^2 \pm x = 0, \quad x^3 = 0$$

and their roots are respectively: 0; $-1, 0, 1$; $-2, -1, -1/2, 0, 1/2, 1, 2$.

The real roots of the equations of higher heights can be obtained likewise, and the set of all real algebraic numbers can be so arranged that it can be counted, i.e. has the cardinal number d .

*14. The set of all real numbers between 0 and 1 is uncountable.

PROOF:

Assume the set to be countable, and list the elements in decimal expansions of the set in a sequence $\{r_1, r_2, r_3, \dots\}$, where they are counted in that order. Since every rational number may be expressed as a repeating decimal and indeed in two ways (e.g. $1/2 = 0.5000\dots = 0.4999\dots$, and in general $0.a_1a_2\dots(a_n+1)000\dots = 0.a_1a_2\dots = a_n999\dots$, ending in an infinite succession of either 0's or 9's), it is agreed not to use the latter type of expansion (with 9's) for the decimals r_1, r_2, \dots in the sequence. List it then in the following array

$$\begin{array}{lcl} r_1 & = & 0.a_{11}a_{12}a_{13}\dots \\ r_2 & = & 0.a_{21}a_{22}a_{23}\dots \\ r_3 & = & 0.a_{31}a_{32}a_{33}\dots \\ & & \dots\dots\dots \end{array}$$

where a_{ij} represents one of 10 digits. Now construct a number $x = 0.b_1b_2b_3\dots$, where b_n is 8 if $a_{nn} \neq 8$ and 7 if $a_{nn} = 8$ (or in any other dichotomy, e.g. 1 if a_{nn} is one of even digits, and 2 if a_{nn} is one of odd digits, etc.). Then, despite the fact that x obviously lies between 0 and 1, it does not belong to the original sequence, since it differs from r_1 in at least the first decimal place, from r_2 in at least the second decimal place, and so on. Hence the initial assumption turns out to be untenable, proving that the given set is not countable.

Generalize this proof, and it can be proved that the set of all real numbers in any interval, and eventually of all real numbers as a whole, is *a fortiori* uncountable; this process, however, involves operations on sets (cf. Df. 2.3.1).

Chapter 2.2

Operations

§2.2.1 Operations in General

Df. 2.2.1.1 An *operation* on a set S is a *code*, i.e. a set of rules, laws, and principles in terms of definitions, metatheorems and theorems, which assign to each ordered subset of n elements of S a uniquely determined element of S .

Example:

The familiar operative rules of addition and multiplication, denoted by “+” and “•”, on the set N of natural numbers where, in this particular case, each ordered subset is always only two elements of N .

The operation in general, then, may be analyzed as follows:

Df. 2.2.1.1a A mathematical operation involves at least three elements:

- (i) an *operand*, the entity which has to be transformed;
- (ii) an *operator*, which symbolizes a rule of manipulation specifying the process of transformation;
- (iii) the *transform*, i.e. the result of the manipulation.

Example:

In the three equations:

$$a + b = c \quad D_x(x^n) = nx^{n-1} \quad \int (\cos x) dx = -\sin x + C$$

the operands are a and b , x^n , $\cos x$, respectively; the operators are $+$, $D_x()$, $\int()dx$, respectively; the transforms are c , nx^{n-1} , $-\sin x + C$, respectively.

An operand, defined as above, is a mathematical entity which as such must be an element of a certain set. Although a set, to be a set, may not need a well-defined operation for itself, an operand cannot be considered without presuming the existence of a certain set; a vacuous operand or an empty operation in a mathematical vacuum is trivial, if not downright meaningless. Hence the following nomenclature:

Df. 2.2.1.1b A set of elements is called a *groupoid* (or any other suitable name) if the elements are considered the operands of a well-defined operation.

In the following pages an operation will always presume a set or sets which can be operated on.

Df. 2.2.1.2 The operation on S is called *unary*, *binary*, *ternary*, ..., in general *n-ary*, depending on the number of each ordered subset of n elements of S to be operated on.

Example:

Squaring, or taking the factorial, or taking the predecessor or successor of a natural number is a unary operation. Integral operation (addition, subtraction, and multiplication) and rational operation (integral operation and division) on the set of integers are binary.

Because it occurs most frequently, the binary operation is re-defined as follows:

Df. 2.2.1.2a A *binary operation*, denoted by $*$ (or \circ or any other suitable symbol), on a set S is a code which assigns to each ordered pair $a, b \in S$ a unique element $c = a * b \in S$.

Example:

The binary operation of *addition*, denoted by $+$, on the set N of natural numbers assigns to each ordered pair, say, n_1 and n_2 , which belong to N , a uniquely determined element $n_3 = n_1 + n_2$, which again belongs to N .

Df. 2.2.1.3 The set S of Df. 2.2.1.2a is said to be *closed* under $*$; in general, a set X , which is a subset of a set Y (i.e. $X \subseteq Y$) is closed under a binary operation \circ , defined on Y , if $a \circ b = c \in X$ for every $a, b \in X$.

Example:

$N \subset I$, where N represents the set of all natural numbers and I the set of all integers, is closed under, say, the binary operation of multiplication, denoted by \cdot , since the operation is defined on I and, for every $a, b \in N$, it is uniquely determined that $a \cdot b = c$, which belongs to N , i.e. $c \in N$.

Df. 2.2.1.4 The binary operation $*$ on a set S is *associative* if, for every $a, b, c \in S$,

$$a * (b * c) = (a * b) * c$$

Example:

The binary operation of addition on the set N of all natural numbers, viz.

$$a + (b + c) = (a + b) + c \quad \text{for every } a, b, c \in N$$

Df. 2.2.1.5 The binary operation $*$ on a set S is *commutative* if, for every $a, b \in S$,

$$a * b = b * a$$

Example:

Addition and multiplication on the set of all natural numbers are both commutative, since $a + b = b + a$ and $a \cdot b = b \cdot a$ for every $a, b \in N$. (On the other hand, subtraction on the set I of all integers is neither associative nor commutative, since it is not always the case that $a - (b - c) = (a - b) - c$ or $a - b = b - a$ for every $a, b, c \in I$.)

MTh. 2.2.1.6 (Principle of Duality). A properly worded valid proposition concerning a certain pair of sets X and Y or operators $*$ and \circ may yield a second valid proposition through the interchange of X and Y or $*$ and \circ ; the first proposition is called the *dual* of the second, and conversely.

Example:

$A = B$, where A and B are two sets, is the dual of $B = A$ (cf. MTh. 2.1.4). Or, in a geometry of two dimensions, the following two propositions are duals: "Any two distinct points on the same plane determine a unique line" and "Any two distinct lines on the same plane determine a unique point"; note how the terms "point" and "line" are interchangeable in the two valid propositions. Likewise, in a geometry of three dimensions, "Any two distinct planes on the same point determine a unique line" and "Any two distinct lines on the same point determine a unique plane" are duals where "line" and "plane" are interchangeable.

Df. 2.2.1.7 Given two binary operations $*$ and \circ on a set S , the operation $*$ is *distributive* under \circ if

$$a * (b \circ c) = (a * b) \circ (a * c),$$

and the operation is *distributive* under $*$ if

$$a \circ (b * c) = (a \circ b) * (a \circ c)$$

If both distributions hold simultaneously, the former is the dual of the latter, and conversely.

Example:

$a \cdot (b + c) = a \cdot b + a \cdot c$ for every $a, b, c \in I$; i.e. multiplication on the set I of all integers is distributive with respect to addition, but addition on I is not distributive with respect to multiplication, since it is not always the case that $a + (b \cdot c) = (a + b) \cdot (a + c)$ for every $a, b, c \in I$. In other algebraic structures, however, both distributions may hold simultaneously (cf. Df. 2.4.1.13 and Df. 2.4.2.1).

Df. 2.2.1.8 An element, denoted by e , of a set S is an *identity* for the binary operation $*$ on S if $a * e = e * a = a$ for every $a \in S$.

Example:

0 of the set I of all integers is the identity for the binary operation of addition on I , since $0 + a = a + 0 = a$ for every $a \in I$; likewise 1 is the identity of the binary operation of multiplication on I , since $1 \cdot a = a \cdot 1 = a$ for every $a \in I$.

Df. 2.2.1.9 If a set S contains an identity e for the binary operation $*$, and if $a * b = b * a = e$ for every $a, b \in S$, then a is an *inverse* of b , and b is an *inverse* of a , in S .

Example:

$a + (-a) = (-a) + a = 0$ for every $a, -a \in I$ implies that a is the inverse of $-a$, and conversely, in I under $+$; likewise, a is the inverse of $1/a$, and conversely, in I under \cdot since $a \cdot (1/a) = (1/a) \cdot a = 1$, $a \neq 0$.

* * * * *

MTh. 2.2.1.10 (Well-Ordering Principle). Every non-vacuous subset of natural numbers contains one, and only one, smallest element.

Example:

$N_1 = \{1, 2, 3, \dots, 100\}$, which is a non-empty subset of the set N of all natural numbers, has the smallest element 1; likewise a non-empty subset $N_2 = \{100, 101, 102\}$, $N_2 \subset N$, has 100 for its smallest element.

This disarmingly simple-looking principle is in fact one of the most fundamental metatheorems upon which other metatheorems can be founded or systematically coordinated (cf. Df. 2.4.1.18, MTh. 2.4.1.19-20).

(Note, also, that this principle, generalized and proved first by Zermelo, is directly related to the critical question in modern mathematics of admissible mathematical methods and mathematical existence problems, which, however, is far beyond the scope of this book.)

On the strength of this metatheorem the following metatheorem, for instance, may be proved (cf. §2.2.2, Prob. 1).

MTh. 2.2.1.11 (Principle of Finite Induction). If S is any non-vacuous subset of natural numbers containing 1 and the integer $n + 1$ for every integer n , $n \in S$, then S contains every natural number.

Stated otherwise: let a proposition $P(n)$ correspond with a positive integer n ; then $P(n)$ is true for all n if, for each positive integer m , the assumption that $P(k)$ is true for all positive integers k , $k \leq m$, implies that $P(m)$ itself is true.

This is in fact the abstract formulation of Mathematical Induction.

§2.2.2 Transformations

Df. 2.2.2.1 A *relation* is represented by a set R , each element of which is an *ordered pair*, denoted by (x, y) , where x is the first *coordinate*, and y the second, of the pair; the term “ordered” specifies the *sense* that $(x, y) \neq (y, x)$ unless the relation is *symmetric* and that $(x, y) = (u, v)$ iff $x = u$ and $y = v$.

Example:

R denoting a relation “is predecessor of”, (x, y) designates “ x is y ’s predecessor” and is obviously different from (y, x) in the same context which designates “ y is x ’s predecessor”; on the other hand, if R denotes a symmetric relation “is equal to”, it then follows that $(x, y) = (y, x)$, since “ x is equal to y ” and “ y is equal to x ” are logically equivalent.

Df. 2.2.2.2 The set R whose elements are the ordered pairs (x, y) may be represented by xRy (which is not the same as yRx unless R is symmetric); and iff xRy , x is said to be *R-related* to y , in which case x is called the *referent*, and y the *relatum*, of R .

Example:

R denoting a relation “is greater than”, xRy designates “ x is greater than y ”, but never “ y is greater than x ” in the same context, where x the greater is the referent and the other the relatum, but not conversely.

Df. 2.2.2.3 Given two sets A, B , and their elements $a \in A, b \in B$, the set C of all pairs (a, b) is called the *Cartesian* (or *direct*) *product*, denoted by $A \times B$, where “ \times ” designates the operator of the product; i.e. $C = A \times B = \{(a, b): (a \in A) \times (b \in B)\}$, which reads “for each pair of a and b , a is an element of A and b an element of B ”.

Example:

If A and B consist of the points on the X and Y axes respectively, representing two sets of all real numbers, then the Cartesian product $C = A \times B$ is the Cartesian plane itself, each point of which is represented by a pair of two real numbers: (x, y) .

Df. 2.2.2.4 A *transformation* is a set T , where T is a (non-empty) subset of the set R of relations, such that no two elements of T have the same first coordinates.

Stated otherwise: a set T is a transformation iff $(x, y) \in T$ and $(x, z) \in T$ imply $y = z$.

T defines thus nothing more than a correspondence between two sets X and Y , merely specifying each element of X to be related to an element of (all or some of) Y ; nevertheless, transformation and *function* may be considered synonymous if the former is ramified (cf. Df. 2.2.2.8 below) as follows:

Df. 2.2.2.5 If X and Y are any two sets, then a *transformation* T (or *mapping* M) of X into or onto Y is an operative rule which assigns to each element x of X a uniquely determined element y of Y ; notationally, $T: X \rightarrow Y$ or $y = T(x)$, where $T(x)$, i.e. y itself, is called the *image* (or *map*) of x by T .

T , then, defines what the usual functional notation f defines, with which the reader is quite familiar. Hence the following definition:

Df. 2.2.2.6 The correspondence T above is also called a *function* f of X into or onto Y ; $y \in Y$, which uniquely corresponds to $x \in X$ by f , is called the *value* of f at x . In the same context the set X is called the *domain* of f and the set Y the *range* of f .

Note. A transformation T of X may map several elements of X onto one and the same element of Y ; e.g. $y = \sin x$, where $x = x' + 2k\pi$ for any integer k , maps all real numbers x onto the same real number $\sin x'$, since $\sin(x' + 2k\pi) = \sin x'$. Note also that each element of Y is not always the image of some element of X ; e.g. if $y > 1$ or $y < -1$, there is then no real number x such that $y = \sin x$.

Df. 2.2.2.7 Two transformations T_1 and T_2 are equal iff $T_1(x) = T_2(x)$ for every $x \in X$.

Example:

If $X = \{1, 2, 3, 4\}$ and $Y = \{1, 2\}$ for which S is a transformation that maps an odd number of X onto 1 of Y and an even number of X onto 2 of Y , and if T is a transformation that maps 1 or 3 of X onto 1 of Y and 2 or 4 of X onto 2 of Y , then $S = T$, since

$$\begin{array}{ll} S(1) = S(3) = 1 & S(2) = S(4) = 2 \\ T(1) = T(3) = 1 & T(2) = T(4) = 2 \end{array}$$

Note, however, that the equality of two transformations cannot be taken into consideration unless their domains and ranges are equal; e.g. if S maps $\{1, 2\}$ into $\{3, 5\}$ through $S(1) = 3$, $S(2) = 5$, and if T maps $\{1, 2\}$ into $\{3, 5, 7\}$ through $T(1) = 3$, $T(2) = 5$, then $S \neq T$ despite their apparent equivalence in transformation.

The transformation in general may be analyzed in a manner the operation in general was dissected (cf. Df. 2.2.2.1a), articulating the customary, and sometimes rather ambiguous, terms "into" and "onto" as follows:

Df. 2.2.2.8 A transformation $T: X \rightarrow Y$ is called

- (i) *surjective* (or *onto*) if $T(x) = y$, i.e. if there is at least one $x \in X$ for every $y \in Y$;
- (ii) *injective* (or *into*) if $T(a) = T(b)$ for every $a, b \in X$ implies $a = b$, or, what is the same, $a \neq b$ implies $T(a) \neq T(b)$; i.e. if every element of Y which is a T -image of an element of X at all is a T -image of one, and only one, element of X ;
- (iii) *bijective* (*onto-into* in the sense of *one-to-one*) if both (i) and (ii) hold simultaneously; i.e. for every $x \in X$ and $y \in Y$, $T(x) = T(y)$ implies $x = y$, or, what is the same, $x \neq y$ implies $T(x) \neq T(y)$.

Example:

$y = \sin x$, described as above with respect to Df. 2.2.2.6, is patently surjective, since the mapping certainly does not exclude the possibility that an element of Y (i.e. $-1 \leq y \leq 1$) may be the image of several elements of X (i.e. $-\infty < x < +\infty$).

On the other hand, $y = \log_e x$, or $x = e^y$, represents an injective mapping, since every element of Y (i.e. $-\infty < y \in Y < +\infty$) that is an image of an element of X (i.e. $0 < x \in X < +\infty$) at all is the image of only one element of X .

Another example of injective mapping is the mapping

$$x \rightarrow (x, y) \equiv (x, f(x))$$

where f is an injection of X into $X \times Y$ (cf. Df. 2.2.2.3), which may represent an ordinary mapping in the Cartesian plane.

As is quite evident in the last example, the definition of functions (cf. Df. 2.2.2.5) in the customary text-book of College Algebra or Calculus *usually* describes an injective transformation, although the wording may not be quite the same as here.

Thirdly, the mapping $(x, y) \rightarrow (y, x)$ of $X \times Y$ into and onto $Y \times X$ is bijective. Or, in a more concrete context, $y = \sin^{-1} x$, or $x = \sin y$, with the following restriction, viz.,

$$|x| \leq 1 \quad \text{and} \quad |y| < \pi/2$$

is a bijective mapping, since the mapping is both onto and into under the restricted domain and range. This is indeed pictorially represented by the principal branch of the inverse sine curve, where the elements of X and Y are exhaustively paired such that every element of X (or Y) has neither more nor less than one element of Y (or X) as its counterpart, completing the curve by the strict 1-1 transformation. It must be emphasized that no part of the principal curve is left out here in the process of mapping.

Df. 2.2.2.9 If X and Y are two sets having an *abstract structure* built on certain operative principles and rules, of the same type, then the transformation $T: X \rightarrow Y$ which preserves the initially defined operations is called a *homomorphism*.

Stated otherwise, in terms of $*$ and \circ (cf. Df. 2.2.1.7): a homomorphism with respect to $*$ and \circ of a set X onto or into a set Y , denoted by $(X, *, Y, \circ)$ -homomorphism, is a transformation X under $*$ onto or into Y under \circ such that, for every $a, b \in X$ and every $T(a), T(b) \in Y$,

$$T(a * b) = T(a) \circ T(b)$$

Note. “ $(X, *, Y, \circ)$ -homomorphism” is to show explicitly how the operations are initially defined for X and Y ; it is quite possible, however, that the two initial operations, denoted by $*$ and \circ respectively, are identical.

Df. 2.2.2.10 An *endomorphism* of X is a homomorphism $T: X \rightarrow X$.

Df. 2.2.2.11 An *isomorphism* is a bijective homomorphism; notationally, $T: X \leftrightarrow Y$, in contrast with the plain surjective or injective homomorphism, $T: X \rightarrow Y$.

The presence of the double arrow “ \leftrightarrow ” in the isomorphism indicates that the mapping $X \rightarrow Y$ can be reversed here, viz. $Y \rightarrow X$; hence an isomorphism is necessarily a 1-1 transformation.

Df. 2.2.2.12 An *automorphism* of X is a bijective endomorphism; notationally, $T: X \leftrightarrow X$.

These four fundamental concepts of transformations are still too abstract to be exemplified at this early stage, but they will soon reappear in due course, incorporated either in entirety or in part in some specific frameworks of algebraic structures, which will be studied in this and other chapters.

Df. 2.2.2.13 If Y_1 is any subset of Y , then the *inverse image* of Y_1 under a transformation T , denoted by $T^{-1}(Y_1)$, is the set of every $x \in X$ whose image is in Y_1 ; if every $x \in X$ has no image in Y_1 , then obviously $T^{-1}(Y_1) = \emptyset$.

Stated otherwise, the 1-1 transformation S , defined by $S(x) = y$ for $T(y) = x$, for every $x \in X, y \in Y$, is the inverse of T , denoted by T^{-1} .

Example:

$x = g(y) = y - 1$ is the inverse of $y = f(x) = x + 1$, for every $x \in X, y \in Y$, where X and Y are two sets of all real numbers; if, however, $y = f(x) = x^2$, then $x = g(y) = \sqrt{y}$ for every $x \in X_1, y \in Y_1$, where X_1 and Y_1 are subsets of X and Y respectively, viz. of all positive real numbers, including 0.

In both examples, $g(y)$ and $f(x)$ may be replaced by $f^{-1}(y)$ and $g^{-1}(x)$ respectively, since one of them is the inverse of the other.

Df. 2.2.2.14 Given two transformations S and T on a set X , the *product* (or *composite*) of S and T , denoted by TS , is a transformation defined by $TS(x) = T(S(x))$ for all $x \in X$.

The composite (or product) of two transformations in terms of the familiar functional notation is a function of a function; e.g. if f and g represent two transformations on X , viz. $f(x) = x^3$ and $g(x) = \sin x$, then $fg(x) = f(g(x)) = \sin^3 x$, and $gf(x) = g(f(x)) = \sin x^3$, revealing that the product of two transformations is not always commutative.

Th. 2.2.2.15 If S is a 1-1 transformation of X into Y and T a 1-1 transformation of Y into Z , then TS is a 1-1 transformation of X into Z . (Cf. Prob. 7 below.)

Example:

If $x = f(y)$ and $y = g(z)$, then $x = f(g(z)) = F(z)$.

3. If a set X consists of n elements, the set S of all possible mappings of X consists of n^n elements.

PROOF:

If $n = 1$, i.e. $X = \{x_1\}$, then S cannot contain more than 1 mapping, i.e. T_1 defined by $T_1(x_1) = x_1$. If $n = 2$, i.e. $X = \{x_1, x_2\}$, then the following 2^2 distinct transformations, T_1, T_2, T_3, T_4 , defined by

$$\begin{array}{llll} T_1(x_1) = x_1 & T_2(x_1) = x_1 & T_3(x_1) = x_2 & T_4(x_1) = x_2 \\ T_1(x_2) = x_1 & T_2(x_2) = x_2 & T_3(x_2) = x_1 & T_4(x_2) = x_2 \end{array}$$

exhaust the possible mappings.

In general, since each element $x_i \in X$ can be mapped onto any of $x_j \in X$, the number of all possible pairings of (x_i, x_j) with respect to T_k , where each of i, j, k may be repeated as often as possible, is n^n . (Out of n different elements, when each may be repeated as we please, the number of ways in which an arrangement of r elements can be made is n^r . Here $r = n$.)

4. Let f and g be two mappings on the set R^* of all real numbers, defined by $f(x) = x + 1$ and $g(x) = x^3 + 1$, $x \in R^*$; find fg and gf .

Solution:

$$\begin{aligned} fg &= f(g(x)) = f(x^3 + 1) = (x^3 + 1) + 1 = x^3 + 2 \\ gf &= g(f(x)) = g(x + 1) = (x + 1)^3 + 1 = x^3 + 3x^2 + 3x + 2 \end{aligned}$$

5. Let S and T be two transformations from $\{1, 2, 3\}$ to itself, defined by: $S(1) = 1$, $S(2) = 2$, $S(3) = 3$; $T(1) = 3$, $T(2) = 2$, $T(3) = 1$. Find $ST(1)$, $TS(2)$, $ST(3)$.

Solution:

$$ST(1) = S(T(1)) = S(3) = 3. \quad TS(2) = T(2) = 2. \quad ST(3) = S(1) = 1.$$

6. If S is a 1-1 transformation of X into Y and T a 1-1 transformation of Y into Z , then TS is a 1-1 transformation of X into Z .

PROOF:

Let $S(x) = y$, $x \in X$, $y \in Y$, and $T(y) = z$, $y \in Y$, $z \in Z$; then $TS(x) = T(y) = z$.

Also, if $x \neq x'$, $x, x' \in X$, then $S(x) \neq S(x')$ and $TS(x) \neq TS(x')$.

Hence TS is a 1-1 transformation of X into Z .

7. If C is a class (cf. Df. 2.3.9) of all 1-1 mappings on a set S , C is closed (cf. Df. 2.2.1.3) with respect to the binary operation of composition.

PROOF:

If $X \subset C$, $Y \subset C$, and $YX(y) = x$, $YX(z) = x$, for $x, y, z \in S$, then $Y(X(y)) = x = Y(X(z))$. Hence, since $Y \subset C$ is a 1-1 mapping, $X(y) = X(z)$, which in turn implies $y = z$, since $X \subset C$ also is a 1-1 mapping.

Hence YX is also a 1-1 mapping, i.e. $YX \subset C$; C is thus closed with respect to composition.

8. The composition of transformations is associative.

PROOF:

Let R, S, T , be any three transformations on a set E , and $x \in E$; then

$$R(ST)(x) = R(ST(x)) = R(S(T(x)))$$

and

$$(RS)T(x) = RS(T(x)) = R(S(T(x))).$$

$$\text{Hence } R(ST) = (RS)T.$$

9. Prove: $TT^{-1} = T^{-1}T = I$, on a set X .

PROOF:

Let $x \in X$ and $T^{-1}(x) = y$, $y \in X$; then $TT^{-1}(x) = T(T^{-1}(x)) = T(y) = x$, since $T^{-1}(x) = y$ iff $T(y) = x$. Hence $TT^{-1} = I$.

Likewise $T^{-1}T = I$, and $TT^{-1} = T^{-1}T = I$.

10. Prove that, for any 1-1 mapping f and its inverse f^{-1} on a set A ,

$$X \subseteq f^{-1}f(X)$$

where $X \subseteq A$ and, by definition, $f(X) = \{f(x_1), f(x_2), \dots, f(x_n), \dots\}$, $x_i \in X$.

PROOF:

If $x \in X$, then $f(x) \in f(X)$, which implies $f^{-1}f(x) \in f^{-1}f(X)$, i.e. $x \in f^{-1}f(X)$. Hence $X \subseteq f^{-1}f(X)$.

11. If $A_1, A_2 \in A$ and $A_1 \supseteq A_2$, then, for any 1-1 mapping f on a set A ,

$$f(A_1) - f(A_2) \subseteq f(A_1 - A_2)$$

PROOF:

Let $b \in f(A_1) - f(A_2)$, where $a \in A_1$ and $f(a) = b$, i.e. $a \notin A_2$; then $b \in f(A_1 - A_2)$ since $a \in A_1 - A_2$.

Hence $f(A_1) - f(A_2) \subseteq f(A_1 - A_2)$.

12. Prove that $(ST)^{-1} = T^{-1}S^{-1}$, then generalize the result.

PROOF:

Since, from Prob. 9 above, $(ST)(ST)^{-1} = I$ and, by Prob. 8-9,

$$(ST)(T^{-1}S^{-1}) = S(TT^{-1})S^{-1} = SIS^{-1} = SS^{-1} = I$$

it follows that $(ST)(ST)^{-1} = (ST)(T^{-1}S^{-1}) = I$ and, by Prob. 6-7, $(ST)^{-1} = T^{-1}S^{-1}$.

When generalized, the new theorem is of the form

$$(T_1 T_2 \dots T_n)^{-1} = T_n^{-1} \dots T_2^{-1} T_1^{-1}$$

and can be proved likewise, viz.,

$$\begin{aligned} T_n^{-1} \dots T_2^{-1} T_1^{-1} (T_1 T_2 \dots T_n) &= T_n^{-1} \dots T_2^{-1} (T_1^{-1} T_1) T_2 \dots T_n = T_n^{-1} \dots (T_2^{-1} T_2) \dots T_n \\ &= \dots = T_n^{-1} T_n = I = (T_1 T_2 \dots T_n)^{-1} (T_1 T_2 \dots T_n) \end{aligned}$$

13. Establish the law of positive exponents for transformations, defining at the start: $T^0 = I$, $T^1 = T$, and $T^{i+1} = TT^i$.

PROOF:

If $m = n = 1$, then $T^m T^n = T^{m+n}$ obviously holds, since $T^1 T^1 = TT = T^{1+1}$.

If the same holds for $m = j$, $n = k$, then for $m = j+1$, $n = k+1$,

$$T^{j+1} T^{k+1} = T^{j+1} T T^k = T^{(j+1)+1} T T^{k-1} = \dots = T^{(j+1)+k} T T^0 = T^{(j+1)+k+1} I = T^{(j+1)+(k+1)}$$

Hence, by induction, for any natural numbers m and n , $T^m T^n = T^{m+n}$.

Likewise, $(T^m)^n = T^{mn}$.

14. Generalize the exponential law of Prob. 13 to all integers m and n , defining $T^{-k} = (T^k)^{-1}$.

PROOF:

The case for $m = n = 0$ is trivial, and the case for $m > 0$ and $n > 0$ has already been established.

If $m' = -m$, $m > 0$, and $n' = -n$, $n > 0$, then

$$\begin{aligned} T^{m'} T^{n'} &= T^{-m} T^{-n} = (T^m)^{-1} (T^n)^{-1} = (T^n T^m)^{-1} = (T^{n+m})^{-1} \\ &= T^{-(n+m)} = T^{-n-m} = T^{-m-n} = T^{m'+n'} \end{aligned}$$

The case for only m (or n) < 0 can be established likewise.

The generalization of $(T^m)^n = T^{mn}$ can be similarly carried out.

Chapter 2.3

Operations on Sets

Df. 2.3.1 The *join* (or *union* or *logical sum*) of two sets X and Y , denoted by $X \cup Y$ which reads “ X cup (or join) Y ”, is the set of all elements which belong to either X or Y or both.

Example:

If $X = \{a, b, c\}$ and $Y = \{b, c, d\}$, then $X \cup Y = \{a, b, c, d\}$; likewise, if I_1 represents the set of 0 and all positive integers and I_2 the set of all negative integers, then $I_1 \cup I_2 = I$, i.e. the set of all integers.

Df. 2.3.2 The *meet* (or *intersection* or *logical product*) of two sets X and Y , denoted by $X \cap Y$ which reads “ X cap (or meet) Y ”, is the set of all elements which belong to both X and Y .

For the same example as above, $X \cap Y = \{b, c\}$ and $I_1 \cap I_2 = \emptyset$.

Df. 2.3.3 The (relative) *complement* (or *difference*) of X with respect to Y , denoted by $Y - X$ (or X'_Y or X_Y^c), is the set of all elements which belong to Y but not to X . If Y is the universal set itself or defined likewise, the complement of X is considered absolute and as such is denoted by $-X$ (or X' or X^c), designating the set of all elements which do not belong to X .

For the same example of Df. 2.3.1, consider I the universe; then I'_1 (or $-I_1$) is the set of all negative integers and I'_2 (or $-I_2$) the set of 0 and all positive integers. (Since both complements in this example are absolute, there is no place for confusion even without the additional “with respect to I ”.)

The outcome of three operations may be represented either one-dimensionally, where X and Y are shown as two partially overlapping segments of a line (cf. Fig. 2.3a), or two-dimensionally, by Venn diagrams (cf. Fig. 2.3b-d).

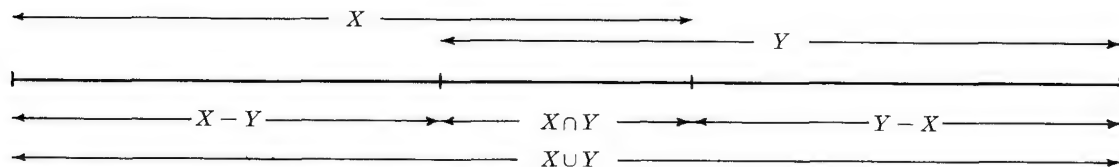
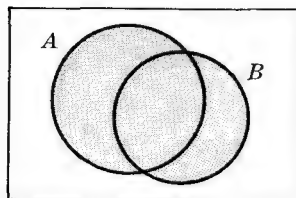
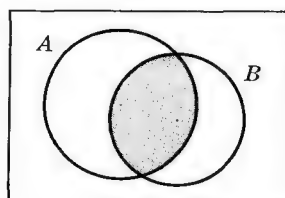


Fig. 2.3a



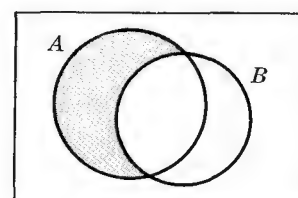
$A \cup B$

Fig. 2.3b



$A \cap B$

Fig. 2.3c



$A - B$

Fig. 2.3d

From the definition itself and figures above there immediately follow the following operative properties with respect to $A - B$:

- (i) If $x \in A$, then $x \in B$ or $x \in (A - B)$.
- (ii) If $x \in B$, then $x \notin (A - B)$.
- (iii) If $x \in (A - B)$, then $x \notin B$.

Note. " $x \in (A - B)$ " may be replaced by " $x \in A - B$ " as long as the operations on sets themselves can be distinguished from those on the elements of sets; sets and elements are said to belong to different *types*.

Df. 2.3.4 If a set S is a subset of U , the universe, then

- (i) $x \in U$ implies that $x \in U$ or $x \in S'$.
- (ii) $x \in S$ implies $x \notin S'$, and $x \in S'$ implies $x \notin S$.
- (iii) $U = \emptyset'$, and $\emptyset = U'$.

Note. Since S is a subset of U , the universe, the complement of S with respect U , viz. $U - S$, is absolute; hence S' (or $-S$).

Df. 2.3.5 Idempotent law:

$$(i) \quad S \cup S = S \qquad (ii) \quad S \cap S = S$$

Df. 2.3.6 Involution law: $(S')' = S$.

Note. If there exist distinct *identity* sets with respect to the operations of join and meet, as in a Boolean Algebra (cf. §2.4.2), then Df. 2.2.2.5-6 become theorems, i.e. can be proved.

Th. 2.3.7 Join and meet are *dually* (cf. MTh. 2.2.1.6) associative, commutative, and distributive (cf. Prob. 7 below):

$$\begin{array}{ll} (i) & A \cup (B \cup C) = (A \cup B) \cup C \qquad A \cap (B \cap C) = (A \cap B) \cap C \\ (ii) & A \cup B = B \cup A \qquad A \cap B = B \cap A \\ (iii) & A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \qquad A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \end{array}$$

Th. 2.3.8 (De Morgan's law). For any three sets A, B, C ,

$$(i) \quad A - (B \cup C) = (A - B) \cap (A - C) \qquad (ii) \quad A - (B \cap C) = (A - B) \cup (A - C)$$

(cf. Prob. 10)

and if A is the universe or otherwise obvious,

$$(i) \quad (B \cup C)' = B' \cap C' \qquad (ii) \quad (B \cap C)' = B' \cup C'$$

The law holds dually and may be extended to any number of sets. (Cf. Prob. 16).

Df. 2.3.9 A *class*, denoted by C (or any capital script letter), is a set of sets. (A class as such belongs to a *type* which is different from the type to which a set or an element of a set belongs.)

Df. 2.3.10 Join and meet may be extended to any number of sets; if C is a class of sets S_i , $i = 1, 2, \dots, n, \dots$, the join of C , denoted by $\cup_{S \in C} S$, is the set of all elements which belong to any of the S_i , and the meet of C , denoted by $\cap_{S \in C} S$, is the set of all elements common in or all of the S_i . How the terms are grouped in computing joins and meets, however, is immaterial because of the following theorem.

Th. 2.3.11 The general associative, commutative, and distributive laws, for $X_i \subset C$, (cf. Prob. 14), are:

(i)-(ii) $\cup_{X \subset C} X$ is unique; so is $\cap_{X \subset C} X$.

(iii) $Y \cup (\cap_{X \subset C} X) = \cap_{X \subset C} (Y \cup X)$; $Y \cap (\cup_{X \subset C} X) = \cup_{X \subset C} (Y \cap X)$.

Df. 2.3.12 Two sets X and Y are said to be *disjoint* iff $X \cap Y = \emptyset$; a class C is disjoint iff no two sets of C intersect, i.e. $\cap_{X \subset C} X = \emptyset$.

Example:

$$I_1 \cap I_2 = \emptyset \quad (\text{cf. Df. 2.3.2}).$$

Df. 2.3.13 A *partition* of a universal set U is a subdivision of U into subsets which are disjoint and exhaustive, called the *cells* of U ; notationally, $U = C_1 \cup C_2 \cup \dots \cup C_n$ where C_i is a cell and $C_i \cap C_j = \emptyset$ if $i \neq j$.

Note. There may be several ways of partitioning one and the same set U . If, e.g., $U = \{a, b, c, d\}$, then $\{\{a\}, \{b, c, d\}\}$, $\{\{a, b\}, \{c, d\}\}$, $\{\{a, c\}, \{b, d\}\}$, \dots , $\{\{a\}, \{b\}, \{c\}, \{d\}\}$, are all distinct partitions of U , the last in particular being the partition of U into its unit sets.

Solved Problems

1. Find X and Y if

(i) $\{1, 2, 3, 5\} - \{1, 3, 8, 9, 10\} = X$

(ii) $\{1, 2, 3\} \cup \{2, 7, 8\} = X$

(iii) $\{2, 3, 4\} \cap \{2, 5, 8\} = X$

(iv) $A - B = X$ and $B - A = Y$ where $A = \{x \mid x \in R^* \geq 0\}$ (which designates “all real numbers greater than, or equal to, 0”), and $B = \{x \mid x \in R^* \leq 0\}$ (which reads “all real numbers less than, or equal to, 0”).

Solution:

(i) $\{2, 5\}$. (ii) $\{1, 2, 3, 7, 8\}$. (iii) $\{2\}$. (iv) $X = \{x \mid x \in R^* > 0\}$, $Y = \{x \mid x \in R^* < 0\}$.

2. If $A \supseteq B$, then $A - C \supseteq B - C$.

PROOF:

If $B - C = \emptyset$, then the proposition is obviously true.

If $B - C \neq \emptyset$, then let $x \in (B - C)$, i.e. $x \in B$ and $x \notin C$. But $x \in A$ ($\because A \supseteq B$), and now $x \notin C$. Hence $x \in (A - C)$. That is, if $x \in (B - C)$, then $x \in (A - C)$, i.e. $A - C \supseteq B - C$.

3. Prove that $A - (A - B) = B$ if $A \supseteq B$; then observe, referring also to Problem 2 above, that Y being the null set in the context of $X \supseteq Y$ is immaterial.

PROOF:

First $A - (A - B) \supseteq B$, then $B \supseteq A - (A - B)$ must be proved.

Let $A - B = C$; if $B = \emptyset$, then evidently $A - C \supseteq B$; and if $B \neq \emptyset$, then, letting $x \in B$, it follows that $x \in A$ and $x \notin (A - B)$, i.e. $x \in A$ and $x \notin C$. Hence $x \in (A - C)$, and $A - C \supseteq B$, i.e. $A - (A - B) \supseteq B$.

Conversely, if $A - C = \emptyset$, evidently $B \supseteq A - C$; and if $A - C \neq \emptyset$, then, letting $x \in (A - C)$, i.e. $x \in A$ and $x \notin C$, viz. $x \in A$ and $x \notin (A - B)$, it follows that $x \in B$. Hence $B \supseteq A - C$, i.e. $B \supseteq A - (A - B)$.

From the proofs of Problems 2, 3, it is obvious that, in the context of $X \supseteq Y$, it is not necessary to examine the case of $Y = \emptyset$ if the case of $Y \neq \emptyset$ is proved valid. This is in fact a truism if it is remembered that the null set is a subset of every set (cf. Df. 2.1.7). Henceforth, in the similar context, only the case of $Y \neq \emptyset$ will be examined.

Note. It can be readily proved likewise that $A \supseteq B$ if $A - (A - B) = B$; hence $A - (A - B) = B$ is in fact the necessary and sufficient condition for $A \supseteq B$. (E.g., in the language of elementary mathematics, if A stands for the set of all real numbers and B for the set of all rational numbers, then $A - B$ is the set of all irrational numbers; take away this set again from the set of all real numbers (i.e. $A - (A - B)$), then what remains is the set B of all rational numbers. Hence $A - (A - B) = B$.)

Note also that the whole proof can be carried out pictorially, i.e. by a Venn-Diagram (cf. Fig. 2.3e).

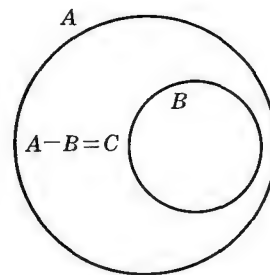


Fig. 2.3e

4. Prove: (ia) $A \subseteq A \cup B$, (ib) $B \subseteq A \cup B$.
 (ii) If $A \subseteq C$ and $B \subseteq C$, then $A \cup B \subseteq C$.
 (iii) If $A \cup B = A$, then $B \subseteq A$, and conversely.

PROOF:

(ia) If $x \in A$, then $x \in A \cup B$, by Df. 2.3.1. Hence $A \subseteq A \cup B$, by Df. 2.1.2. Likewise (ib).

(ii) Let $x \in A \cup B$; then $x \in A$ or $x \in B$. In either case, since $A \subseteq C$ and $B \subseteq C$, it follows that $x \in C$. Hence $A \cup B \subseteq C$.

(iii) Since $B \subseteq A$, and also obviously $A \subseteq A$, it follows from (ii) that $A \cup B \subseteq A$. And, from (i), $A \subseteq A \cup B$. Hence $A \cup B = A$.

Conversely, since $A \cup B = A$, it immediately follows that $A \cup B \subseteq A$, and, from (i), $B \subseteq A \cup B$. Hence, by Th. 2.1.3, $B \subseteq A$.

5. Prove: (ia) $A \cap B \subseteq A$, (ib) $A \cap B \subseteq B$.
 (ii) If $C \subseteq A$ and $C \subseteq B$, then $C \subseteq A \cap B$.
 (iii) If $B \subseteq A$, then $A \cap B = B$, and conversely.

PROOF:

(ia) Since $A \cap B = \{x \mid x \in A \text{ and } x \in B\}$, $x \in A \cap B$ implies $x \in A$. Hence $A \cap B \subseteq A$.
 Likewise (ib).

(ii) If $x \in C$, then $x \in A$ and $x \in B$ ($\because C \subseteq A$ and $C \subseteq B$ by hypothesis). Hence $C \subseteq A \cap B$.

(iii) Since $B \subseteq A$ and, obviously, $B \subseteq B$, it follows from (ii) that $B \subseteq A \cap B$. Also, from (i), $A \cap B \subseteq B$. Hence $A \cap B = B$.

Conversely if $x \in B$, then $x \in A \cap B$ ($\because A \cap B = B$ by hypothesis). But, from (i), $A \cap B \subseteq A$, which implies by Df. 2.1.2 that $x \in A$. Hence $B \subseteq A$.

6. Prove: (i) $A \cup B = A \cap B$ iff $A = B$.
 (ii) $A \cup B = (A - B) \cup B$.
 (iii) $(A - B) \cup B = A$ iff $A \supseteq B$.

PROOF:

- (i) If $A = B$, i.e. $A \supseteq B$ and $A \subseteq B$, then from Prob. 4iii and Prob. 5iii, $A \cup B = A = B = A \cap B$, i.e. $A \cup B = A \cap B$. Conversely, if $A \cup B = A \cap B$, i.e. $A \cup B \supseteq A \cap B$ and $A \cup B \subseteq A \cap B$, then from Prob. 4ia and Prob. 5ia, $A \cap B \subseteq A \subseteq A \cup B$, and from Prob. 4ib and Prob. 5ib, $A \cap B \subseteq B \subseteq A \cup B$, i.e. $A \cup B = A = B = A \cap B$, i.e. $A = B$.
- (ii) $(A - B) \cup B \subseteq A \cup B$, since $A - B \subseteq A \subseteq A \cup B$ and $B \subseteq A \cup B$. Now, if $x \in A \cup B$, then $x \in A$ or $x \in B$. If $x \in B$, then $x \in (A - B) \cup B$, and if $x \notin B$, then $x \in A$ and $x \in A - B$, i.e. $x \in (A - B) \cup B$. Hence $A \cup B \subseteq (A - B) \cup B$, and together with its converse proved at the start, $A \cup B = (A - B) \cup B$.
- (iii) From (ii), $(A - B) \cup B = A \cup B$. Hence $(A - B) \cup B = A$ iff $A = A \cup B$, i.e. $A \supseteq B$ (cf. Prob. 4iii).

7. Prove that join and meet are dually associative, commutative, and distributive (cf. Th.2.2.2.4), then verify it, using the following three sets: $A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$, $C = \{1, 3, 5\}$.

PROOF:

These operative rules will be proved only with respect to join, since the proof with respect to meet will then be similarly and quite easily carried out, due to their duality.

- (i) $A \cup (B \cap C) = \{x \mid x \in A \text{ or } (x \in B \text{ or } x \in C)\} = \{x \mid (x \in A \text{ or } x \in B) \text{ or } x \in C\} = (A \cup B) \cup C$. Likewise $A \cap (B \cup C) = (A \cap B) \cap C$.
- (ii) $A \cup B = \{x \mid x \in A \text{ or } x \in B\} = \{x \mid x \in B \text{ or } x \in A\} = B \cup A$. Likewise $A \cap B = B \cap A$.
- (iii) Let $x \in A \cup (B \cap C)$; then $x \in A$ or $x \in B \cap C$. If $x \in A$, then $x \in A \cup B$ and $x \in A \cup C$ ($\because A \subseteq A \cup B$, $A \subseteq A \cup C$); hence $x \in (A \cup B) \cap (A \cup C)$. Also, if $x \in B \cap C$, then $x \in B$ and $x \in C$, i.e. $x \in A \cup B$ and $x \in A \cup C$; hence $x \in (A \cup B) \cap (A \cup C)$. In either case $A \cup (B \cap C) \subseteq (A \cup B) \cap (A \cup C)$.

Conversely, let $x \in (A \cup B) \cap (A \cup C)$, i.e. $x \in A \cup B$ and $x \in A \cup C$; then $x \in A$ or $x \in B$ and $x \in C$, i.e. $x \in A$ or $x \in B \cap C$. Hence $(A \cup B) \cap (A \cup C) \subseteq A \cup (B \cap C)$. Thus, altogether, $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$. Likewise $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

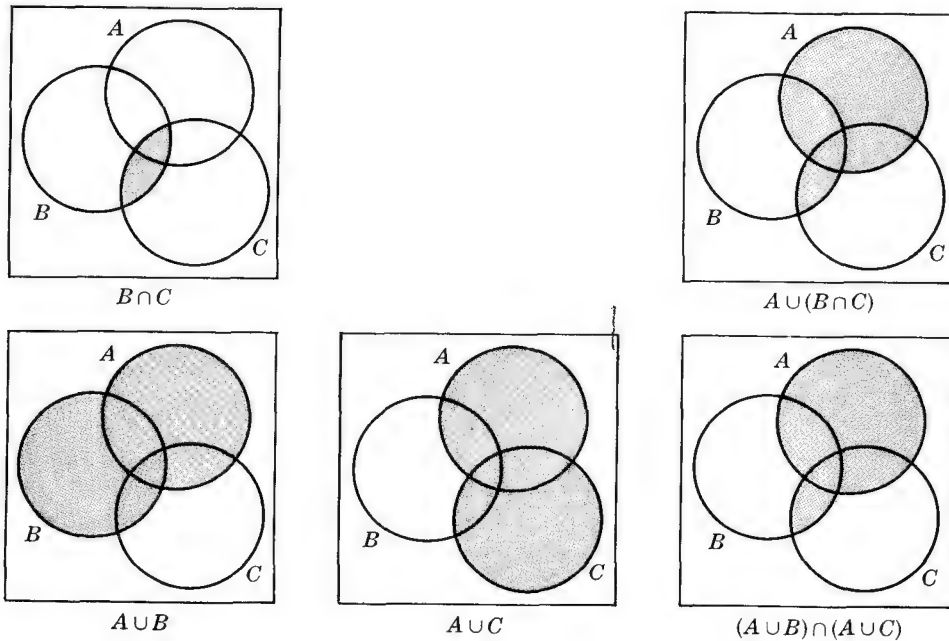
Given $A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$, $C = \{1, 3, 5\}$:

$$A \cup (B \cap C) = \{1, 2, 3, 4, 5\} = (A \cup B) \cup C; \quad A \cap (B \cup C) = \{3\} = (A \cap B) \cap C.$$

$$A \cup B = \{1, 2, 3, 4, 5\} = B \cup A; \quad A \cap B = \{3\} = B \cap A.$$

$$A \cup (B \cap C) = \{1, 2, 3, 5\} = (A \cup B) \cap (A \cup C); \quad A \cap (B \cup C) = \{1, 3\} = (A \cap B) \cup (A \cap C)$$

As a second proof, one of the distributive laws, for instance, can be pictorially verified by Venn diagrams as follows:



8. Referring to §2.2.2, prove that, for any 1-1 mapping of a set A into a set B ,

$$f(A \cap B) \subseteq f(A) \cap f(B)$$

where, by definition, $f(S) = \cup_{x \in S} f(x)$ for any set S .

PROOF:

If $x \in f(A \cap B)$, then, by definition, $x \in f(A)$ and $x \in f(B)$ hold simultaneously, i.e. $x \in f(A) \cap f(B)$.

Hence $f(A \cap B) \subseteq f(A) \cap f(B)$.

9. In the same context as above (Prob. 8), prove that

$$f(A \cap f^{-1}(B)) = f(A) \cap B$$

where $f^{-1}(Y) = \cup_{y \in Y} f^{-1}(y)$, $Y \subseteq B$, when $f(X) = \cup_{x \in X} f(x)$, $x \subseteq A$.

PROOF:

Directly from Problem 8, it follows that $f(A \cap f^{-1}(B)) \subseteq f(A) \cap f(f^{-1}(B)) = f(A) \cap B$, i.e. $f(A \cap f^{-1}(B)) \subseteq f(A) \cap B$. Conversely, if $b \in f(A) \cap B$, then $f(a) = b$ for $a \in A$, i.e. $a \in f^{-1}(B)$; and since $b \in B$, it follows that $a \in A \cap f^{-1}(B)$. Hence $b \in f(A \cap f^{-1}(B))$, and $f(A) \cap B \subseteq f(A \cap f^{-1}(B))$.

Hence, from two conclusions, it follows that $f(A \cap f^{-1}(B)) = f(A) \cap B$.

10. Prove De Morgan's law: (i) $A - (B \cup C) = (A - B) \cap (A - C)$, (ii) $A - (B \cap C) = (A - B) \cup (A - C)$; then verify the law with three sets A, B, C , given as in Problem 7.

PROOF:

Let $x \in A - (B \cup C)$, i.e. $x \in A$ and $x \notin B \cup C$; then $x \notin B$ and $x \notin C$, i.e. $x \in A - B$ and $x \in A - C$, i.e. $x \in (A - B) \cap (A - C)$. Hence $A - (B \cup C) \subseteq (A - B) \cap (A - C)$.

Conversely, if $x \in (A - B) \cap (A - C)$, i.e. $x \in A - B$ and $x \in A - C$, then $x \in A$ and $x \notin B$, $x \notin C$, i.e. $x \in A$ and $x \notin B \cup C$, i.e. $x \in A - (B \cup C)$. Hence $(A - B) \cap (A - C) \subseteq A - (B \cup C)$.

Hence $A - (B \cup C) = (A - B) \cap (A - C)$.

Likewise $A - (B \cap C) = (A - B) \cup (A - C)$.

Given $A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$, $C = \{1, 3, 5\}$:

$$A - (B \cup C) = \{1, 2, 3\} - \{1, 3, 4, 5\} = \{2\} = \{1, 2\} \cap \{2\} = (A - B) \cap (A - C)$$

$$A - (B \cap C) = \{1, 2, 3\} - \{3, 5\} = \{1, 2\} = \{1, 2\} \cup \{2\} = (A - B) \cup (A - C)$$

11. Referring to Df. 2.2.2.3 (on the Cartesian product), prove that

$$(ia) \quad A \times (B \cup C) = (A \times B) \cup (A \times C), \quad (ib) \quad A \times (B \cap C) = (A \times B) \cap (A \times C)$$

$$(iia) \quad (A \cup B) \times C = (A \times C) \cup (B \times C), \quad (iib) \quad (A \cap B) \times C = (A \times C) \cap (B \times C).$$

PROOF:

(ia) Let $(x, y) \in A \times (B \cup C)$, i.e. $x \in A$, $y \in B \cup C$; then $y \in B$ or $y \in C$, i.e. $(x, y) \in A \times B$ or $(x, y) \in A \times C$. Hence $(x, y) \in (A \times B) \cup (A \times C)$, and $A \times (B \cup C) \subseteq (A \times B) \cup (A \times C)$.

Conversely, if $(x, y) \in (A \times B) \cup (A \times C)$, then $y \in B$ and, since $y \in B \cup C$, $(x, y) \in A \times (B \cup C)$.

Hence $(A \times B) \cup (A \times C) \subseteq A \times (B \cup C)$, and thus $A \times (B \cup C) = (A \times B) \cup (A \times C)$.

Likewise (ib), (iia), (iib).

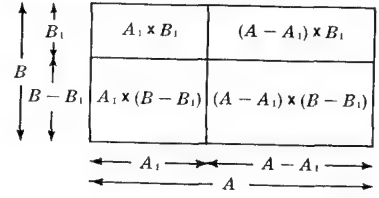
12. Prove that (i) addition of finite cardinal numbers is associative and commutative, and (ii) multiplication of finite cardinal numbers is associative, commutative, and distributive with respect to addition.

PROOF:

- (i) The two laws follow directly from the associativity and commutativity of the join of sets.
- (ii) Let A, B, C be any three sets of finite elements, viz. $o(A) = a$, $o(B) = b$, $o(C) = c$; then, by Df. 2.1.11 and Problem 2 of §2.2.2,
- (iia) $o(A \times (B \times C)) = a(bc) = (ab)c = o((A \times B) \times C)$
- (iib) $o(A \times B) = ab = ba = o(A \times B)$
- (iic) $o(A \times (B \cup C)) = a(b + c) = ab + ac = o((A \times B) \cup (A \times C))$ (cf. Problem 11, ia above).

13. If A and B are any two sets and $A_1 \subseteq A$ and $B_1 \subseteq B$, then

$$\begin{aligned} A \times B - A_1 \times B_1 &= ((A - A_1) \times B) \cup (A_1 \times (B - B_1)) \\ &= ((A - A_1) \times B_1) \cup (A \times (B - B_1)) \\ &= (A_1 \times (B - B_1)) \cup ((A - A_1) \times B_1) \\ &\quad \cup ((A - A_1) \times (B - B_1)) \end{aligned}$$



PROOF:

Cf. Fig. 2.3f, where the proof is completed by partition (cf. Df. 2.3.13).

Fig. 2.3f

14. Now that the Principle of Finite Induction is affirmed (cf. MTh. 2.2.1.11), generalize the three fundamental laws of association, commutation, and distribution with respect to join and meet (cf. Th. 2.3.8).

PROOF:

The case for $n=1$ is trivial for all three laws.

For $n=2$, the commutative law has already been proved (Prob. 7, ii), and the associative law $A \cup (A \cup B) = (A \cup A) \cup B$ or any of its equivalents also holds; so does the distributive law $A \cup (A \cap B) = (A \cup A) \cap (A \cup B)$ or any of its equivalents.

For $n=3$, the associative or distributive law has already been proved (Prob. 7, i, iii), and the commutative law $A \cup (B \cup C) = (B \cup C) \cup A = A \cup B \cup C$ also holds.

Their duals with respect to meet for $n=1, 2, 3$ are also true.

In general, assume that the associative law with respect to join (or meet) holds for $n=k$, viz.,

$$\begin{aligned} S_1 \cup (S_2 \cup (\dots (S_{k-1} \cup S_k) \dots)) &= (S_1 \cup S_2) \cup (S_3 \cup (\dots (S_{k-1} \cup S_k) \dots)) = \dots \\ &= (\dots ((S_1 \cup S_2) \dots) \cup S_{k-1}) \cup S_k = S_1 \cup S_2 \cup \dots \cup S_{k-1} \cup S_k \end{aligned}$$

Then, for $n = k+1$,

$$\begin{aligned} S_1 \cup (S_2 \cup (\dots S_{k-1} \cup (S_k \cup S_{k+1})) \dots) &= (\dots ((S_1 \cup S_2) \dots) \cup S_{k-1}) \cup (S_k \cup S_{k+1}) \\ &= S_1 \cup S_2 \cup \dots \cup S_{k-1} \cup S_k \cup S_{k+1}, \end{aligned}$$

and the law must generally hold.

The commutative law for n sets is proved likewise.

As for the distributive law, assume that it holds for $n=k$, viz.,

$$S_1 \cup (S_2 \cap \dots \cap S_{k-1} \cap S_k) = (S_1 \cup S_2) \cap \dots \cap (S_1 \cup S_{k-1}) \cap (S_1 \cup S_k)$$

Then, for $n = k+1$, viz.,

$$\begin{aligned} S_1 \cup (S_2 \cap \dots \cap S_k \cap S_{k+1}) &= S_1 \cup ((S_2 \cap \dots \cap S_k) \cap S_{k+1}) \\ &= (S_1 \cup (S_2 \cap \dots \cap S_k)) \cap (S_1 \cup S_{k+1}) = (S_1 \cup S_2) \cap \dots \cap (S_1 \cup S_{k+1}) \end{aligned}$$

Hence the law generally holds.

The general duals with respect to meet can be proved likewise.

15. Prove that, for any 1-1 mapping f on a class C of the sets S_i , $i = 1, 2, \dots, n$,

$$f(\cup_{S \in C} S) = \cup_{S \in C} (f(S))$$

PROOF:

If $b \in f(\cup_{S \in C} S)$, then there exists a such that $a \in \cup_{S \in C} S$ where $b = f(a)$. This implies that $a \in S_i$ for any $S_i \in C$, i.e. $b = f(a) \in f(S_i)$. Hence $f(\cup_{S \in C} S) \subseteq \cup_{S \in C} (f(S))$.

Conversely, going backwards, $\cup_{S \in C} (f(S)) \subseteq f(\cup_{S \in C} S)$. Thus, together, $f(\cup_{S \in C} S) = \cup_{S \in C} (f(S))$.

16. Generalize De Morgan's law, viz.

$$(i) \quad A - \cup_{S \in C} S = \cap_{S \in C} (A - S), \quad (ii) \quad A - \cap_{S \in C} S = \cup_{S \in C} (A - S),$$

Or more simply, considering A the universe,

$$(i) \quad (\cup_{S \in C} S)' = \cap_{S \in C} S', \quad (ii) \quad (\cap_{S \in C} S)' = \cup_{S \in C} S'.$$

PROOF:

$$\begin{aligned} A - \cup_{S \in C} S &= \{x \mid x \in A \text{ and } x \notin \cup_{S \in C} S\} = \{x \mid x \in A \text{ and } x \notin S_i, S_i \in C, i = 1, 2, \dots, n, \dots\} \\ &= \{x \mid x \in (A - S_i)\} = x \in \cap_{S \in C} (A - S). \text{ Hence } A - \cup_{S \in C} S = \cap_{S \in C} (A - S). \end{aligned}$$

Likewise for (ii).

Second proof. Use the Induction Principle. For $n=1$ the case is trivial, and for $n=2$ the validity of the case has already been established by Prob. 8 above.

In general, assume that the law holds for $n=k$, viz.,

$$A - (S_1 \cup S_2 \cup \dots \cup S_k) = (A - S_1) \cap (A - S_2) \cap \dots \cap (A - S_k)$$

Then, for $n = k+1$,

$$\begin{aligned} A - \cup_{S \in C} S &= A - (S_1 \cup \dots \cup S_k \cup S_{k+1}) = A - ((S_1 \cup \dots \cup S_k) \cup S_{k+1}) \\ &= (A - (S_1 \cup \dots \cup S_k)) \cap (A - S_{k+1}), \quad \text{by De Morgan's law itself for } n=2 \\ &= (A - S_1) \cap \dots \cap (A - S_k) \cap (A - S_{k+1}) = \cap_{S \in C} (A - S). \end{aligned}$$

The duals are proved likewise.

17. If $S_0 \in C$, then $\cap_{S \in C} S \subseteq S_0 \subseteq \cup_{S \in C} S$.

PROOF:

Let $x \in \cap_{S \in C} S$; then, by hypothesis, $x \in S_0$, and $\cap_{S \in C} S \subseteq S_0$. And if $x \in S_0$, then $x \in \cup_{S \in C} S$; hence $S_0 \subseteq \cup_{S \in C} S$.

If $S_0 = S_1 = \dots = S_i = \dots$, then obviously $\cap_{S \in C} S = S_0 = \cup_{S \in C} S$.

Hence, in general, $\cap_{S \in C} S \subseteq S_0 \subseteq \cup_{S \in C} S$.

18. Deduce the duals of the distributive law of join and meet, using the involution law (cf. Df. 2.3.6) and De Morgan's law.

PROOF:

$$\begin{aligned} A \cap (B \cup C) &= ((A \cap (B \cup C))')' = (A' \cup (B \cup C)')' = (A' \cup (B' \cap C'))' = ((A' \cup B') \cap (A' \cup C'))' \\ &= (A' \cup B')' \cup (A' \cup C')' = (A'' \cap B'') \cup (A'' \cap C'') = (A \cap B) \cup (A \cap C). \end{aligned}$$

Its dual is deduced likewise.

19. Prove $A \cap (A \cup B) = A$; then deduce from itself its dual, i.e. $A \cup (A \cap B) = A$, as above, using the involution law and De Morgan's law.

PROOF:

(i) Obviously $A \cap (A \cup B) \subseteq A$. On the other hand, from $A \subseteq A$ and $A \subseteq A \cup B$, it follows that $A \subseteq A \cap (A \cup B)$. Hence $A \cap (A \cup B) = A$.

(ii) $A \cup (A \cap B) = ((A \cup (A \cap B))')' = (A' \cap (A \cap B)')' = (A' \cap (A' \cup B'))' = (A')' = A$.

20. Find S' of $S = S_1 \cup S_2 \cap S_3 \cup ((S_4 \cup S_5) \cap (S_6 \cup S_7 \cap S_8))$.

Solution:

$$\begin{aligned} S' &= (S_1 \cup S_2 \cap S_3 \cup ((S_4 \cup S_5) \cap (S_6 \cup S_7 \cap S_8)))' \\ &= S_1' \cap (S_2 \cup S_3)' \cup ((S_4 \cup S_5) \cap (S_6 \cup (S_7 \cap S_8)))' \\ &= S_1' \cap (S_2' \cup S_3') \cap ((S_4 \cup S_5)' \cup (S_6 \cup (S_7 \cap S_8))') \\ &= S_1' \cap (S_2' \cup S_3') \cap ((S_4' \cap S_5') \cup (S_6' \cap (S_7' \cap S_8')')) \\ &= S_1' \cap (S_2' \cup S_3') \cap ((S_4' \cap S_5') \cup (S_6' \cap (S_7' \cap S_8'))). \end{aligned}$$

21. $A - B = A$ iff A and B are disjoint, i.e. $A \cap B = \emptyset$.

PROOF:

Since $A \cap B = \emptyset$, i.e. $\{x \mid x \in A \text{ and } x \notin B\}$, if $x \in A - B$, then $x \in A$, i.e. $A - B \subseteq A$; also if $x \in A$, then $x \in A - B$, i.e. $A \subseteq A - B$. Hence $A \cap B = \emptyset$ implies $A - B = A$.

Conversely, if $A - B = A$, i.e. $\{x \mid x \in A \text{ and } x \notin B\} = \{x \mid x \in A\}$, then obviously $A \cap B = \emptyset$.

22. If two sets A and B , where $A \cap B = \emptyset$, are countable, so is $C = A \cup B$.

PROOF:

- (i) If both A and B are finite, i.e. $o(A) = m$ and $o(B) = n$, then $o(C) = m + n$, which is evidently countable.
- (ii) If $o(A) = m$ and $o(B) = d$ in the sense of Df. 2.1.14 (or likewise $o(A) = d$ and $o(B) = n$), then let $A = \{a_1, a_2, \dots, a_m\}$ and $B = \{b_1, b_2, \dots, b_n, \dots\}$, and arrange the elements of $A \cup B$ so as to be countable as follows:

$$C = \{a_1, a_2, \dots, a_m, b_1, b_2, \dots, b_n\}$$

- (iii) If both A and B are infinite and countable, viz. $o(A) = o(B) = d$, then arrange the elements of C in such a countable manner as follows:

$$C = \{a_1, b_1, a_2, b_2, \dots, a_n, b_n, \dots\}$$

Hence C is countable in all three cases, completing the proof.

23. The join of a countable number of mutually disjoint countable sets is also countable.

PROOF:

Let $S_i = \{a_{i1}, a_{i2}, \dots, a_{ij}, \dots\}$, and arrange $S_1, S_2, \dots, S_i, \dots$ as follows:

$$\begin{aligned} S_1 &= a_{11}, a_{12}, a_{13}, a_{14}, \dots \\ S_2 &= a_{21}, a_{22}, a_{23}, a_{24}, \dots \\ &\dots\dots\dots \\ S_i &= a_{i1}, a_{i2}, a_{i3}, a_{i4}, \dots \\ &\dots\dots\dots \end{aligned}$$

Their join is then countable in several ways (cf. §2.1, Problem 12).

24. Establish the existence of uncountably many real transcendental numbers (i.e. real numbers which are not algebraic), using Problems 13-14 of §2.1.

PROOF:

If the set T of all real transcendental numbers were countable, then the join of T and A , the set of all real algebraic numbers which is countable (cf. Prob. 13 of §2.1), also must be countable. Hence the set \bar{R} of all real numbers ($= A \cup T$, by definition) also must be countable; that is, the set \bar{R}_0 of all real numbers between 0 and 1 must be *a fortiori* countable, which is contrary to what Problem 14 of §2.1 proved. Hence T cannot be denumerated.

Chapter 2.4

Abstract Structures

*§2.4.1 Lattices

Df. 2.4.1.1 A *binary relation* on a set S is a set R of propositional functions such that if $x, y \in S$, then either xRy (reading “ x stands in the relation R to y ”) or $\neg xRy$ (i.e. negation of xRy). In either case x is called the *referent* and y the *relatum* of R (cf. Df. 2.2.2.1).

Df. 2.4.1.2 R on S is *reflexive* if xRx holds, *antisymmetric* when xRy and yRx hold simultaneously iff $x = y$, and *transitive* when xRz if xRy and yRz , where $x, y, z \in S$ (cf. Df. 2.1.12).

Df. 2.4.1.3 S is *partly ordered* with respect to R if R on S is reflexive, antisymmetric, and transitive.

Example:

The set N of all natural numbers under a binary relation, \leq (or \geq , cf. Df. 4.1.2.12-13), is partly ordered, since, for $x, y, z \in N$: (i) $x \leq x$ (or $x \geq x$); (ii) $x \leq y$ and $y \leq x$ (or $x \geq y$ and $y \geq x$) imply $x = y$; (iii) $x \leq y$ and $y \leq z$ (or $x \geq y$ and $y \geq z$) imply $x \leq z$ (or $x \geq z$). Other examples are: “is contained in”, “is subset of”, (both denoted by “ \subseteq ”), etc.

Df. 2.4.1.4 Two partly ordered sets S_1 and S_2 are *isomorphic* (cf. Df. 2.2.2.11) if there exists a 1-1 correspondence T between S_1 and S_2 such that, for $x \in S_1$ and $y \in S_2$,

$$T(x) \supseteq T(y) \text{ iff } x \supseteq y$$

Th. 2.4.1.5 Any set which is partly ordered with respect to R (cf. Df. 2.4.1.1) is likewise partly ordered with respect to the dual \bar{R} of R (cf. Problem 1).

Stated otherwise: The *converse* of any partial ordering is itself a partial ordering. (\bar{R} is the “converse” of R such that $x\bar{R}y$ (reading “ x is in the relation \bar{R} to y ”) iff yRx ; e.g. \bar{R} of “is contained in” is “contains”).

Df. 2.4.1.6 If $x, y \in S$ and both xRp and yRp hold for $p \in S$, then p is called an *upper bound* of x and y ; if pRq holds for any upper bound $q \in S$ of x and y , then p is the *join* or *least upper bound* (or, abbreviated, *l.u.b.* or *sup*, an abbreviation of supremum) of x and y .

Example:

10 is an upper bound of the set consisting of the numbers $-7, -1, 5, 8$, and 9 ; so is the number 11 of the open intervals $(-9, 5)$ and $(7, 11)$ and also of the closed intervals $[-13, -11]$ and $[9, 11]$. There cannot be a number x that is an upper bound of the set of all positive integers, or of all positive real numbers for that matter, for x is evidently less than the number $x + 1$.

Whenever a set is bounded above, it may have many upper bounds, since an upper bound x implies many other upper bounds, $x + 1$, etc., of which x may be the l.u.b. if there exists no upper bound less than x itself.

Df. 2.4.1.7 Dually, if $x, y \in S$ and both pRx and pRy hold for $p \in S$, then p is called a *lower bound* of x and y ; if qRp holds for any lower bound $q \in S$ of x and y , then p is the *meet* or *greatest lower bound* (*g.l.b.* or *inf*, an abbreviation of infimum) of x and y .

The examples of lower bounds and the g.l.b. are readily constructed in parallel to those of upper bounds and the l.u.b.

Df. 2.4.1.8 A *lattice* is a partly ordered set L ($L \subset S$), any two of whose elements x and y have a join, denoted by $x \cup y$, and a meet, denoted by $x \cap y$.

Example:

The set R^* of all real numbers under " \leq " (meaning as usual "is less than or equal to"), etc. (Cf. Prob. 4-6.)

Df. 2.4.1.9 If there exists a 1-1 mapping T of L_1 into L_2 such that, for $x, y \in L_1$ and $T(x), T(y) \in L_2$,

$$T(x \cup y) = T(x) \cup T(y) \quad \text{and} \quad T(x \cap y) = T(x) \cap T(y)$$

then L_1 and L_2 are said to be *isomorphic*.

Th. 2.4.1.10 The operators of L , viz. join and meet, are interchangeable in any theorem with respect to L . (Cf. Prob. 2.)

Th. 2.4.1.11 The commutative, associative, absorption, and idempotent laws hold in L (cf. Prob. 7), viz.,

$$\mathbf{L1.} \quad x \cup y = y \cup x \quad \text{and} \quad x \cap y = y \cap x$$

$$\mathbf{L2.} \quad x \cup (y \cap z) = (x \cup y) \cap z \quad \text{and} \quad x \cap (y \cup z) = (x \cap y) \cup z$$

$$\mathbf{L3.} \quad x \cup (x \cap y) = x \quad \text{and} \quad x \cap (x \cup y) = x$$

$$\mathbf{L4.} \quad x \cup x = x \quad \text{and} \quad x \cap x = x$$

Th. 2.4.1.12 The four laws L1-4 of Th. 2.4.1.11 completely characterize L . (Cf. Prob. 8.)

Df. 2.4.1.13 If L satisfies the distributive law,

$$\mathbf{L5.} \quad x \cup (y \cap z) = (x \cup y) \cap (x \cup z)$$

(and the dual, cf. Prob. 11), L is then called a *distributive lattice*.

Df. 2.4.1.14 If for every $x \in L$ there exists an x' such that

$$x \cup x' = u \quad \text{and} \quad x \cap x' = u'$$

(where u and u' are called the *universal bounds*), L is then called a *complemented lattice*.

Df. 2.4.1.15 L is called *modular* (or *Dedekind*) iff $x \supseteq z$ implies $x \cap (y \cup z) = (x \cup y) \cap z$.

Th. 2.4.1.16 If L is distributive, it is then modular. (Cf. Prob. 14.)

Df. 2.4.1.17 If L is distributive and complemented, it is called a *Boolean lattice*.

* * * * *

A review of the Well-ordering Principle (cf. Df. 2.2.1.10) in relation to partial ordering may help understanding both concepts.

Df. 2.4.1.18 The *ordering* in general may satisfy some of the following axioms: Given a set S whose elements are a, b, c, \dots ,

$$\mathbf{O1.} \quad a \leq b \text{ and } b \leq a \text{ imply } a = b.$$

$$\mathbf{O2.} \quad a \leq b \text{ and } b \leq c \text{ imply } a \leq c.$$

$$\mathbf{O3.} \quad \text{Either } a \leq b \text{ or } b \leq a \text{ for any } a, b \in S.$$

$$\mathbf{O4.} \quad \text{Any non-empty subset } R \text{ of } S \text{ has an element } r_1 \text{ such that } r_1 \leq r \text{ for any } r \in R.$$

The *partial* ordering satisfies **O1, O2**, the *simple* ordering **O1, O2, O3**, and the *well* ordering **O1, O2, O3, O4**. The Well-ordering Principle is logically equivalent to the following metatheorems:

MTh. 2.4.1.19 (Axiom of Choice). If C is a class of disjoint non-empty sets $S_i, i = 1, 2, \dots$, then there exists a set S which consists of exactly one element x_i each from S_i .

MTh. 2.4.1.20 (Zorn's Lemma). If every simply ordered subset of a partially ordered set S has an upper bound (or a lower bound), then S has at least one maximal (or minimal) element.

Solved Problems

1. Prove Th. 2.4.1.5.

PROOF:

Since the dual \bar{R} is characterized by

$$(i) \quad x\bar{R}x \quad (ii) \quad y = x \text{ iff } y\bar{R}x \text{ and } x\bar{R}y \quad (iii) \quad z\bar{R}x \text{ if } z\bar{R}y \text{ and } y\bar{R}x$$

R and \bar{R} are evidently isomorphic, i.e.,

$$(i) \quad xRx \leftrightarrow x\bar{R}x \quad (ii) \quad x = y \text{ iff } xRy \text{ and } yRx \leftrightarrow y = x \text{ iff } y\bar{R}x \text{ and } x\bar{R}y \\ (iii) \quad xRz \text{ if } xRy \text{ and } yRz \leftrightarrow z\bar{R}x \text{ if } z\bar{R}y \text{ and } y\bar{R}x$$

Hence S remains partly ordered under \bar{R} .

(The proof will be verified *in concreto* if R and \bar{R} are replaced by " \leq " and " \geq ".)

2. Prove Th. 2.4.1.10.

PROOF:

From Th. 2.4.1.5 it directly follows that, L being *a fortiori* a partly ordered set,

$$x \cup y, x, y \in L \text{ under } R \leftrightarrow x \cap y, x, y \in L \text{ under } \bar{R}$$

Hence the join may be replaced by the meet, and conversely, in any theorem with respect to L .

Note. This theorem validates the dualization of all theorems with respect to L .

3. If a set S is a partly ordered set which has $x \cup y$ (or $x \cap y$), where $x, y \in S$, then the join (or meet) is unique.

PROOF:

Suppose S has two joins, j_1 and j_2 , for x and y ; then, since $j_1, j_2 \in S$, there exist $j_1 R_{j_2}$ and $j_2 R_{j_1}$, from which it follows, by Df. 2.4.1.3, that $j_1 = j_2$. Hence the join is unique.

The case of the dual can be proved likewise.

4. Prove that the set R^* of all real numbers under " \leq " (or " \geq ") forms a lattice.

PROOF:

R^* is partly ordered, satisfying Df. 2.4.1.3, and if $x, y \in R^*$, then $x \cup y$ is the greater (or the equal) of x and y , and $x \cap y$ is the smaller (or the equal) of x and y .

Hence R^* under \leq is a lattice.

So is R^* under \geq , as can be proved likewise.

5. The set N of all natural numbers under “ $|$ ”, designating integral division (i.e. “ $x|y$ ” meaning “ x divides y ”, x being an exact divisor of y), forms a lattice.

PROOF:

N is *a fortiori* partly ordered, satisfying Df. 2.4.1.3; also, $x \cup y$ is here the least common multiple of x and y , and $x \cap y$ is their highest common factor.

Hence N under “ $|$ ” is a lattice.

6. If K is a class consisting of all subsets of a set S , then K is a lattice.

PROOF:

K being a class whose elements are sets, it has the relation of inclusion, by Df. 2.1.2, and is partly ordered. Also, K has joins and meets for any $X, Y \in K$, exactly the way defined by Df. 2.3.1-2, viz. $X \cup Y$ and $X \cap Y$. Hence K is a lattice.

7. Establish the commutative, associative, absorption, and idempotent laws for L under R .

PROOF:

- (i) **The Commutative law:** Let $x \cup y = p$ and $y \cup x = q$ for $x, y, p, q \in L$; then either pRq or qRp . If pRq , then $(x \cup y)R(y \cup x)$, and if qRp , then $(y \cup x)R(x \cup y)$. Hence, L being a partly ordered set and by Df. 2.4.1.3, it follows from $(x \cup y)R(y \cup x)$ and $(y \cup x)R(x \cup y)$ that

$$x \cup y = y \cup x$$

The dual can be proved likewise, viz. $x \cap y = y \cap x$.

- (ii) **The Associative law:** Let $x \cup (y \cup z) = p$ for $x, y, z, p \in L$; then, by Df. 2.4.1.6, xRp and $(y \cup z)Rp$. But then, by the same Df., $yR(y \cup z)$ and $zR(y \cup z)$; hence, from $yR(y \cup z)$ and $(y \cup z)Rp$, it follows that, by Df. 2.4.1.3, yRp , and likewise zRp . Hence xRp, yRp, zRp , i.e. p is an upper bound of x, y, z .

Let q be any upper bound of x, y, z ; then, by Df. 2.4.1.6, xRq and $(y \cup z)Rq$; and applying the same Df. again, $x \cup (y \cup z)Rq$, i.e. pRq . Hence p is the l.u.b. of x, y, z .

On the other hand, let $(x \cup y) \cup z = r$ for $x, y, z \in L$; then, going through the same steps as above, it follows that rRs for any upper bound s of x, y, z , and r is the l.u.b. of x, y, z .

But then, by Problem 3, $p = r$, i.e.,

$$x \cup (y \cup z) = (x \cup y) \cup z$$

The dual can be proved likewise, viz. $x \cap (y \cap z) = (x \cap y) \cap z$.

- (iii) **The Absorption law:** By Df. 2.4.1.7, $(x \cap y)Rx$ and, from Df. 2.4.1.3, xRx ; hence x is an upper bound of $x \cap y$ and x . If w is any upper bound of these two, it is then an upper bound of any of the two, e.g. xRw . Hence x is the l.u.b. of x and $x \cap y$, i.e.,

$$x \cup (x \cap y) = x$$

The dual can be proved likewise, viz. $x \cap (x \cup y) = x$.

- (iv) **The Idempotent law:** Applying (iii) to $x \cup x$,

$$x \cup x = x \cup (x \cap (x \cup y)) = x \cup (x \cap Y) = x$$

where Y is a substitute for $x \cup y$ (or any element, for that matter). Hence

$$x \cup x = x$$

The dual can be proved likewise, viz. $x \cap x = x$.

8. The four laws L1-4 of Th. 2, 4, 1.10 (cf. Problem 7 above) completely characterize L .

PROOF:

In any abstract structure which satisfies L1-4, $x \cup y = y$ iff $x \cap y = x \cap (x \cup y) = x$, and by defining $x \leq y$ to mean $x \cup y = y$, the structure becomes a lattice where $x \cup y$ is the l.u.b. of x and y and $x \cap y$ the g.l.b. of x and y .

E.g. (i) $x \cup x = x$ entails the reflexive property of L . (ii) $y = x \cup y = y \cup x = x$, by L1, if $x \leq y$ and $y \leq x$, affirming the antisymmetric nature of L . (iii) $x \cup z = x \cup (y \cup z) = (x \cup y) \cup z = y \cup z = z$, by L2, if $x \leq y$ and $y \leq z$, justifying the transitivity of L .

Furthermore, since $x \cup (x \cup y) = (x \cup x) \cup y = x \cup y$, by L1, 4, $x \cup y$ is an upper bound of x and, by L1, also of y ; but it is the l.u.b. of x and y , since $x \leq z$ and $y \leq z$ imply $(x \cup y) \cup z = x \cup (y \cup z) = x \cup z = z$.

Also, dually, $x \cap y$ is the g.l.b. of x and y , which completes the proof.

9. If $x, y, z \in L$, then $x \cup (y \cap z) \leq (x \cup y) \cap (x \cup z)$ and $x \cap (y \cup z) \geq (x \cap y) \cup (x \cap z)$.

PROOF:

Since, by Df. 2.4.1.7, $x \geq x \cap y$ and $x \geq x \cap z$, it follows that $x \geq (x \cap y) \cup (x \cap z)$. Also $y \cup z \geq (x \cap y) \cup (x \cap z)$, since $(y \cup z) \geq y \geq (x \cap y)$ and $(y \cup z) \geq z \geq (x \cap z)$.

Hence $x \cup (y \cap z) \geq (x \cap y) \cup (x \cap z)$.

The dual can be proved likewise.

10. For every $a, b, c, d \in L$, (i) $a \cap (b \cup c) \geq (a \cap b) \cup c$ if $a \geq c$, and
(ii) $a \cup c \geq b \cup d$ and $a \cap c \geq b \cap d$ if $a \geq b$ and $c \geq d$.

PROOF:

(i) Since $a \cap c = c$, (i) follows directly from Problem 9.

(ii) Since $a \cup c \geq b$ and $a \cup c \geq d$, by Df. 2.4.1.6, it immediately follows that $a \cup c \leq b \cup d$. The dual can be proved likewise.

11. If a distributive law is given as L5 in L , e.g. $x \cup (y \cap z) = (x \cup y) \cap (x \cup z)$, then the second distributive law, viz. $x \cap (y \cup z) = (x \cap y) \cup (x \cap z)$, can be deduced as the dual from L5.

PROOF:

Applying L5, 1, 4, 3, 2, 1, 4 successively,

$$\begin{aligned} (x \cap y) \cup (x \cap z) &= ((x \cap y) \cup x) \cap ((x \cap y) \cup z) = (x \cup (x \cap y)) \cap (z \cup (x \cap y)) \\ &= x \cap ((z \cup x) \cap (z \cup y)) = (x \cap (z \cup x)) \cap (z \cup y) \\ &= (x \cap (x \cup z)) \cap (y \cup z) = x \cap (y \cup z) \end{aligned}$$

12. L is modular iff, for every $x, x', y \in L$, $x \cup y = x' \cup y$, $x \cap y = x' \cap y$, and $x \geq x'$ imply $x = x'$.

PROOF:

(i) If L is modular, then $x = x'$ under the given conditions, since $x \cap (y \cup x') = (x \cup y) \cap x'$, by Df. 1.4.1.15, where $x \cap (y \cup x') = x \cap (x' \cup y) = x \cap (x \cup y) = x$, by L1, 3, and $(x \cup y) \cap x' = x' \cap (x \cup y) = x' \cap (x' \cup y) = x'$, by L1, 3.

(ii) If $x = x'$, then L is modular. For, if L is not modular, there exist x, x', y such that $x \geq x'$ and $x' \cup (y \cap x) \neq (x' \cup y) \cap x$. Now $x' \cup (y \cap x) < (x' \cup y) \cap x$, since $x' \leq x' \cup y$ and $y \cap x \leq x$. Then, letting $x' \cup (y \cap x) = X$ and $(x' \cup y) \cap x = Y$,

$$X \cup y \leq Y \cup y \leq (x' \cup y) \cup y = x' \cup y = x' \cup (y \cup (y \cap x)) = (x' \cup (y \cap x)) \cup y = X \cup y$$

proving that $X \cup y = Y \cup y$ for $X < Y$, which is a contradiction. Hence L must be modular if $x = x'$.

13. L is a distributive lattice iff

$$(x \cup y) \cap (y \cup z) \cap (z \cup x) = (x \cap y) \cup (y \cap z) \cup (z \cap x) \quad \text{for } x, y, z \in L$$

PROOF:

(i) If L is distributive, then

$$\begin{aligned} (x \cup y) \cap (y \cup z) \cap (z \cup x) &= (((x \cup y) \cap (y \cup z)) \cap z) \cup (((x \cup y) \cap (y \cup z)) \cap x) \\ &= ((x \cup y) \cap x) \cup ((y \cup z) \cap x) = ((x \cap z) \cup (y \cap z)) \cup ((y \cap x) \cup (z \cap x)) \\ &= (x \cap y) \cup (y \cap z) \cup (z \cap x) \end{aligned}$$

(ii) Conversely, if (i) holds, then

$$\begin{aligned} x \cap ((x \cap y) \cup (y \cap z) \cup (z \cap x)) &= (x \cap y) \cup x \cap ((y \cap z) \cup (z \cap x)) \quad (\because x \cup y \leq x) \\ &= (x \cap y) \cup ((z \cap x) \cup (x \cap (y \cap z))) \quad (\because z \cap x \leq x) \\ &= (x \cap y) \cup (z \cap x) \end{aligned}$$

On the other hand,

$$x \cap (x \cup y) \cap (y \cup z) \cap (z \cup x) = (x \cap (x \cup y)) \cap ((z \cup x) \cap (y \cup z)) = (x \cap (x \cup z)) \cap (y \cup z) = x \cap (y \cup z)$$

Hence $x \cap (y \cup z) = (x \cap y) \cup (x \cap z)$, i.e. distributive.

14. If L is distributive, it is then modular.

PROOF:

Let $x \leq z$; then $x \cup (y \cap z) = (x \cup y) \cap (x \cup z) = (x \cup y) \cap z$, which, by Df. 2.4.1.15, completes the proof.

15. \mathbf{K} of Prob. 6 is a Boolean lattice.

PROOF:

If $X, Y, Z \subseteq \mathbf{K}$, then each element of \mathbf{K} can be complemented and there also exists a dual distributive law with respect to $X, Y, Z \subseteq \mathbf{K}$.

16. Prove that any algebraic structure A which for any $x, y, z \in A$, satisfies

A1. $x \cup x = x$

A2. $x \cup u = u \cup x = u$ for a universal bound $u \in A$

A3. $x \cap u = u \cap x = x$

A4. $x \cap (y \cup z) = (x \cap y) \cup (x \cap z)$ and $(y \cup z) \cap x = (y \cap x) \cup (z \cap x)$

is a distributive lattice with a universal bound u .

PROOF:

(i) **L4** is complete if the dual of **A1** exists, viz.,

$$\begin{aligned} x &= x \cap u = x \cap (x \cup u) = (x \cap x) \cup (x \cap u) = (x \cap x) \cup (x \cap (x \cup u)) \\ &= (x \cap x) \cup ((x \cap x) \cup (x \cap x)) = (x \cap x) \cup (x \cap x) \cup u \\ &= (x \cap x) \cap (x \cap x) = x \cap x \end{aligned}$$

A1a

(ii) **L3** is immediately provided as follows:

$$x \cup (x \cap y) = (x \cap u) \cup (x \cap y) = x \cap (u \cup y) = x \cap u = x$$

A5a

and similarly

$$(x \cap y) \cup x = x$$

A5b

Then, using **A1**, 4, 5a,

$$x \cap (x \cup y) = (x \cap x) \cup (x \cap y) = x \cup (x \cap y) = x \quad \text{A5c}$$

and similarly

$$x \cap (y \cup x) = x \quad \text{A5d}$$

Note that **A5a** and **A5c** are equivalent to **L3**.

(iii) Applying **A5c-d**, 4,

$$\begin{aligned} x \cup y &= (x \cap (y \cup x)) \cup (y \cap (x \cup y)) = (x \cup y) \cap (y \cup x) \\ &= ((x \cup y) \cap y) \cup ((x \cup y) \cap x) = y \cup x \end{aligned} \quad \text{A6a}$$

which proves one-half of **L1**.

(iv) Applying **A4**, 5a, 5c,

$$x \cap ((x \cup y) \cup z) = (x \cap (x \cup y)) \cup (x \cap z) = x \cup (x \cap z) = x \quad \text{A7a}$$

and similarly

$$y \cap ((x \cup y) \cup z) = y \quad \text{A7b}$$

and

$$z \cap ((x \cup y) \cup z) = z \quad \text{A7c}$$

Applying **A7a-c**, **A4** twice, and **A1**,

$$\begin{aligned} x \cup (y \cup z) &= (x \cap ((x \cup y) \cup z)) \cup (y \cap ((x \cup y) \cup z)) \cup (z \cap ((x \cup y) \cup z)) \\ &= ((x \cup y) \cap ((x \cup y) \cup z)) \cup (z \cap ((x \cup y) \cup z)) \\ &= ((x \cup y) \cup z) \cap ((x \cup y) \cup z) = (x \cup y) \cup z \end{aligned} \quad \text{A8a}$$

which proves one-half of **L2**.

(v) Applying **A8a**, 5a,

$$\begin{aligned} (x \cup y) \cap (x \cup z) &= (x \cap (x \cup z)) \cup (y \cap (x \cup z)) = x \cup ((y \cap x) \cup (y \cap z)) \\ &= (x \cup (y \cap x)) \cup (y \cap z) = x \cup (y \cap z) \end{aligned} \quad \text{A4a}$$

which proves the rest of **A4**, now completing **L5**, and similarly

$$(x \cap y) \cup z = (x \cup z) \cap (y \cup z) \quad \text{A4b}$$

(vi) Since **A4a-b** is now available, the duals of **A6a** and **A8a** can be proved likewise, thus completing **L1-5**.

§2.4.2 Boolean Algebras

Df. 2.4.2.1 A *Boolean algebra* is a set B of elements x, y, z, \dots , operated on by the dual binary operators of *join* (union or disjunction) or *meet* (or intersection or conjunction), denoted by \vee and \wedge respectively, as follows:

- B1.** Closure: (1a) $x \vee y \in B$ if $x, y \in B$ (1b) $x \wedge y \in B$ if $x, y \in B$
B2. Commutative law: (2a) $x \vee y = y \vee x$ (2b) $x \wedge y = y \wedge x$
B3. Distributive law: (3a) $x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z)$ (3b) $x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$
B4. Identity: (4a) $x \vee O = x$ (4b) $x \wedge I = x$
 (O and I correspond to \emptyset and U of Df. 2.1.6-7, called here the *zero element* and the *universal element* respectively.)
B5. Complement: (5a) $x \vee x' = I$ (5b) $x \wedge x' = O$
B6. Inclusion: $x \subset y$ (reading “ x is included in y ”) iff $x' \vee y = I$

B contains other fundamental properties such as the laws of association, idempotence, involution, absorption, etc., which, however, can be deduced from **B1-6** (cf. Prob. 1-11).

Boolean algebras are special lattices, viz.:

Th. 2.4.2.2 A Boolean lattice is a Boolean algebra; i.e. if a lattice L is distributive and complemented, it becomes a Boolean algebra B through corresponding joins and meets, e.g.

$$x \cup y \leftrightarrow x \vee y \quad \text{and} \quad x \cap y \leftrightarrow x \wedge y \quad (\text{Cf. Prob. 22})$$

Th. 2.4.2.3 Conversely, B becomes L under the relation of inclusion, **B6**, denoted by \subset , with corresponding joins and meets (cf. Prob. 23).

Note that **B6** can be defined in various ways, viz.:

Th. 2.4.2.4 $x \subset y$ iff $x \vee y = y$ or iff $x \wedge y = x$ or iff $x \wedge y' = O$. (Cf. Prob. 12-16.)

(This theorem also may be considered exemplifying the difference between mathematical and everyday languages (cf. Df. 1.1.1.6 and notes); for, if the theorem is to be interpreted in ordinary language, it is patently false, since a single “if and only if” in the latter demands the exclusion of all but one “iff”, in which the theorem above abounds.)

Boolean algebras as abstract mathematical *structures* may become *models* for other mathematical or logical or even industrial systems, e.g. as follows:

Df. 2.4.2.5 There exists a 1-1 correspondence between a binary Boolean algebra B and an algebra C of circuit designs as follows:

B	\vee	\wedge	$-$	U	\emptyset
C	\updownarrow	\updownarrow	\updownarrow	\updownarrow	\updownarrow
	parallel	series	negation	on	off
	circuit	circuit			

(Cf. Prob. 24-26 below.)

Df. 2.4.2.6 There exists a 1-1 correspondence between a binary Boolean algebra B and a two-valued algebra P of propositions under the connectives of “or”, “and”, and “not” as follows:

B	\vee	\wedge	$-$ (or $'$)	U	\emptyset
P	\updownarrow	\updownarrow	\updownarrow	\updownarrow	\updownarrow
	and/or	and	not	true	false

Th. 2.4.2.7 The algebra P of propositions is a Boolean algebra. (Cf. Prob. 27 below.)

The outcome of the theorem is that PART I of this book, in particular §1.1.1, Tautologies, is nothing but a Boolean algebra, and that what can be asserted on the strength of the definitions, metatheorems, and theorems of §1.1.1 can be asserted likewise by Df. 2.4.2.1 and its theorems, and conversely.

Solved Problems

- Both identity elements, viz. O and I , of B are unique.

PROOF:

Suppose there exist two zero elements, O_1 and O_2 where $O_1 \neq O_2$; then, by **B4a**,

$$O_1 \vee O_2 = O_1 \quad \text{and} \quad O_2 \vee O_1 = O_2$$

But, by **B2a**,

$$O_1 \vee O_2 = O_2 \vee O_1$$

Hence $O_1 = O_2$, i.e. O is unique.

The uniqueness of I can be proved likewise.

- Any complement, viz. x' for x , in B is unique.

PROOF:

Let x'_1 and x'_2 be both complements of x in B such that $x \vee x'_1 = I$, $x \vee x'_2 = I$, $x \wedge x'_1 = O$, and $x \wedge x'_2 = O$. But, then, applying **B4b**, **3b**, **2a**, **2a**, **4a** successively,

$$\begin{aligned} x'_1 &= x'_1 \wedge I \equiv x'_1 \wedge (x \vee x'_2) = (x'_1 \wedge x) \vee (x'_1 \wedge x'_2) \\ &= (x \wedge x'_1) \vee (x'_1 \wedge x'_2) \equiv O \vee (x'_1 \wedge x'_2) = (x'_1 \wedge x'_2) \vee O = x'_1 \wedge x'_2 \end{aligned}$$

And, taking exactly the same steps, $x'_2 = x'_1 \wedge x'_2$, i.e. $x'_1 = x'_2$. Hence x' of x is always unique.

Note. The notation " \equiv " in the proof is to show the steps where only the principle of substitution, and not any of **B1-6**, is applied. This distinction is necessary whenever such a difference is substantial.

- Each of the identity elements in B is the complement of the other, viz. $O' = I$ and $I' = O$.

PROOF:

There exists O' for O in B and, as Problem 2 has proved, O' is unique. Now, by **B4a**, $O' \vee O = O'$ and, by **B5a**, $O \vee O' = I$. But, by **B2a**, $O' \vee O = O \vee O'$. Hence $O' = I$.

$I' = O$ can be proved likewise.

- Prove the idempotent law for B , viz. (a) $x \vee x = x$ and (b) $x \wedge x = x$.

PROOF:

Applying **B4a**, **5b**, **3a**, **5a**, **4b** successively,

$$x = x \vee O = x \vee (x \wedge x') = (x \vee x) \wedge (x \vee x') = (x \vee x) \wedge I = x \vee x$$

The dual can be proved likewise.

5. Prove the involution law for B , viz. $(x')' \equiv x'' = x$.

PROOF:

Since x' is unique, by Problem 2, so is x'' ; and again by Problem 2 itself, $x'' \vee O = x''$, $x'' \wedge I = x''$, $x' \vee x'' = I$, $x' \wedge x'' = O$.

Now suppose $x'' \neq x$; then, since $x'' = x'' \vee O$ and, applying **B5b, 3a, 4a** step by step,

$$x'' \vee O = x'' \vee (x \wedge x') = (x'' \vee x) \wedge (x'' \vee x') = x'' \vee x, \quad \text{i.e. } x'' = x'' \vee x$$

and since $x'' \neq x$, it follows from Problem 4 that $x'' \vee x \neq x''$, i.e. $x'' \neq x''$, which, however, is contradictory to the Principle of Identity (cf. MTh. 2.1.1a). Hence $x'' = x$.

Note. The same can be obtained by starting with $x'' \wedge I$ and $x'' \neq x$.

6. Prove: (a) $x \vee I = I$, (b) $x \wedge O = O$.

PROOF:

Applying **B4b, 2b, 5a, 4b, 5a** successively,

$$x \vee I = (x \vee I) \wedge I = I \wedge (x \vee I) = (x \vee x') \wedge (x \vee I) = x \vee (x' \wedge I) = x \vee x' = I$$

Likewise, applying **B5a, 2a, 5b, 3a, 4a, 5a**,

$$x \wedge O = (x \wedge O) \vee O = O \vee (x \wedge O) = (x \wedge x') \vee (x \wedge O) = x \wedge (x' \vee O) = x \wedge x' = O$$

7. Prove the absorption law for B , viz.

(a) $x \vee (x \wedge y) = x$ and (b) $x \wedge (x \vee y) = x$; then prove that $x = y$ if $x \vee z = y \vee z$ and $x \wedge z = y \wedge z$.

PROOF:

Applying **B2a, 4a, 3a, Prob. 6a, B4a** successively

$$x \wedge (x \vee y) = (x \vee y) \wedge x = (x \vee y) \wedge (x \vee O) = x \vee (y \wedge O) = x \vee O = x$$

The dual can be proved likewise. Furthermore,

$$x = x \wedge (x \vee z) = x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z) = (x \wedge y) \vee (y \wedge z) = y \wedge (x \vee z) = y \wedge (y \vee z) = y$$

i.e. $x = y$, by repeated application of the laws of absorption and substitution.

8. If $a = c$ and $b = d$, then $a \vee b = c \vee d$.

PROOF:

By MTh. 2.1.1a, $X = X$; then if $X \equiv x \vee y$, $x \vee y = x \vee y$. Let $x \equiv a$ and $y \equiv b$; then, by MTh. 1.1.1.9, $a \vee b = a \vee b$. But, by hypothesis, $a = c$ and $b = d$, which implies, by MTh. 1.1.1.9 again, $a \vee b = c \vee d$, completing the proof.

9. Prove the associative law for B , viz.

(a) $x \vee (y \vee z) = (x \vee y) \vee z$ and (b) $x \wedge (y \wedge z) = (x \wedge y) \wedge z$.

PROOF:

(a) Let $X \equiv x \vee (y \vee z)$ and $Y \equiv (x \vee y) \vee z$; then, applying **B3b, Prob. 4b-7b**,

$$x \wedge X = (x \wedge x) \vee (x \wedge (y \vee z)) = x \vee (x \wedge (y \vee z)) = x$$

and applying **B3a, Prob. 7a-b**,

$$x \wedge Y = (x \wedge (x \vee y)) \vee (x \wedge z) = x \vee (x \wedge z) = x$$

Hence $x \wedge X = x \wedge Y$.

On the other hand, applying **B3b**, **2b**, **5b**, **2a**, **3a** successively,

$$\begin{aligned} x' \wedge X &= (x' \wedge x) \vee (x' \wedge (y \vee z)) = (x \wedge x') \vee (x' \wedge (y \vee z)) \\ &= O \vee (x' \wedge (y \vee z)) = (x' \wedge (y \vee z)) \vee O = x' \wedge (y \vee z) \end{aligned}$$

and applying **B3b**, **3b**, **2a**, **5a**, **2b**, **4a**, **3a**,

$$\begin{aligned} y' \wedge X &= (x' \wedge (x \vee y)) \vee (x' \wedge z) = ((x' \wedge x) \vee (x' \wedge y)) \vee (x' \wedge z) \\ &= ((x \wedge x') \vee (x' \wedge y)) \vee (x' \wedge z) = (O \vee (x' \wedge y)) \vee (x' \wedge z) \\ &= ((x' \wedge x) \vee O) \vee (x' \wedge z) = (x' \wedge y) \vee (x' \wedge z) = x' \wedge (y \vee z) \end{aligned}$$

Hence $x' \wedge X = y' \wedge X$.

Applying Prob. 8 here, let $a = x \wedge X$, $b = x' \wedge X$, $c = x \wedge Y$, $d = y' \wedge X$, and $(x \wedge X) \vee (x' \wedge X) = (x \wedge Y) \vee (y' \wedge X)$. Then, applying **B2b**, **3a**, **5a**, **4b**, step by step,

$$\begin{aligned} (X \wedge x) \vee (X \wedge x') &= (Y \wedge x) \vee (Y \wedge x'), \\ \text{i.e. } X \wedge (x \vee x') &= Y \wedge (x \vee x'), \quad \text{i.e. } X \wedge I = Y \wedge I, \quad \text{i.e. } X = Y \end{aligned}$$

Hence $x \vee (y \vee z) = (x \vee y) \vee z$.

(b) Likewise, $x \wedge (y \wedge z) = (x \wedge y) \wedge z$.

10. Prove De Morgan's law for B , viz. (a) $(x \vee y)' = x' \wedge y'$ and (b) $(x \wedge y)' = x' \vee y'$.

PROOF:

(a) $(x \vee y)' = x' \wedge y'$ obviously holds if it is proved that

$$(x \vee y) \vee (x' \wedge y') = I \quad \text{and} \quad (x \vee y) \wedge (x' \wedge y') = O \quad (\text{cf. B5 and Problem 2})$$

Applying **B3a**, **2a**, Prob. 9a, **B2a**, **5a**, Prob. 6a, **B4b** successively,

$$\begin{aligned} (x \vee y) \vee (x' \wedge y') &= ((x \vee y) \vee x') \wedge ((x \vee y) \vee y') = (x' \vee (x \vee y)) \wedge ((x \vee y) \vee y') \\ &= ((x' \vee x) \vee y) \wedge (x \vee (y \vee y')) = (y \vee I) \wedge (x \vee I) = I \wedge I = I \end{aligned}$$

Likewise, applying **B2b**, **3b**, **2b**, Prob. 9b, **B2b**, **5b**, Prob. 6b, **B4a**,

$$\begin{aligned} (x \vee y) \wedge (x' \wedge y') &= (x' \wedge y') \wedge (x \vee y) = ((x' \wedge y') \wedge x) \vee ((x' \wedge y') \wedge y) \\ &= (x \wedge (x' \wedge y')) \vee ((x' \wedge y') \wedge y) = ((x \wedge x') \wedge y') \vee (x' \wedge (y' \wedge y)) \\ &= (y' \wedge (x \wedge x')) \vee (x' \wedge (y' \wedge y)) = (y' \wedge O) \vee (x' \wedge O) = O \vee O = O \end{aligned}$$

(b) The dual can be proved likewise.

11. Prove that inclusion in B (cf. **B6**) is dual, i.e. $x \subset y$ and $x \supset y$ are dual.

PROOF:

The dual of " $x \subset y$ iff $x' \vee y = 1$ " is " $x \supset y$ iff $x' \wedge y = O$ ", which is true, since

$$x' \vee y = O \leftrightarrow (x' \wedge y)' = O' = 1 \leftrightarrow (x')' \vee y' = 1 \leftrightarrow x \vee y' = 1 \leftrightarrow y \subset x \leftrightarrow x \supset y$$

Note that this theorem validates the "dualization" of all other inclusion theorems. Note, also, that the double arrow signs are employed here, as in Prob. 13-20 below, to show the logical equivalence between any two successive steps in the proof.

12. Prove that $x \subset y$ iff $x \vee y = y$.

PROOF:

If $x \vee y = y$, then $x' \vee (x \vee y) = (x' \vee x) \vee y = (x \vee x') \vee y = 1 \vee y = y$, i.e. $x' \vee (x \vee y) = x' \vee y = 1$. Hence $x \subset y$.

Conversely, if $x \subset y$, i.e. $x' \vee y = 1$, then $x \vee y = (x \vee y) \wedge 1 = (x \vee y) \wedge (x' \vee y) = (y \vee x) \wedge (y \vee x') = y \vee (x \wedge x') = y \vee O = y$, i.e. $x \vee y = y$.

13. Prove: $x \subset x$.

PROOF:

By Prob. 4a, $x \vee x = x$, which is, by Prob. 12, logically equivalent to $x \subset x$, completing the proof.

Second proof. $x \subset x \leftrightarrow x' \vee x = x \vee x' = 1$, by B2a, 4a, 6.

14. $x \subset y$ iff $x \wedge y' = O$.

PROOF:

$$x \subset y \leftrightarrow x' \vee y = I \leftrightarrow (x' \vee y)' = I' \leftrightarrow (x')' \wedge y' = O \leftrightarrow x \wedge y' = O, \text{ by B6, Prob. 3, 10, 5.}$$

15. $x \subset y$ iff $x \wedge y = x$.

PROOF:

If $x \wedge y = x$, then, applying Prob. 9b, B5b, Prob. 6b,

$$x \wedge y' = (x \wedge y) \wedge y' = x \wedge (y \wedge y') = x \wedge O = O$$

i.e. $x \wedge y' = O$; hence, by Prob. 14, $x \subset y$.

Conversely, if $x \subset y$, i.e. $x \wedge y' = O$, by Prob. 14, then by applying B4a, Prob. 14, B3a, 5a, 4a,

$$x \wedge y = (x \wedge y) \vee O = (x \wedge y) \vee (x \wedge y') = x \wedge (y \vee y') = x \wedge I = x$$

16. $x \wedge y = x$ iff $x \vee y = y$ (or iff $x \wedge y' = O$).

PROOF:

$$x \wedge y = x \leftrightarrow x \subset y \leftrightarrow x \vee y = y \quad (\leftrightarrow x \wedge y' = O), \text{ by Prob. 12, 15, (14).}$$

17. If $x \subset y$ and $y \subset z$, then $x \subset z$.

PROOF:

Since $x \vee y = y$ and $y \vee z = z$, by Problem 16,

$$x \vee z = x \vee (y \vee z) = (x \vee y) \vee z = y \vee z = z,$$

by substitutions and Problem 9a, i.e. $x \vee z = z$. Hence, by Problem 16 again, $x \subset z$.

Second method: Since $x \wedge y = x$, by Problem 15, and $y \vee z = z$, by Problem 16,

$$x \wedge z = x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z) = x \vee (x \wedge z) = x$$

Hence, by Problem 15 again, $x \subset z$. (Other methods can be similarly devised.)

18. If $x \in B$, then $O \subset x \subset I$.

PROOF:

Since $O \vee x = x \vee O = x$, by B2a, 4a, i.e. $O \vee x = x$, it directly follows from Problem 12 that $O \subset x$; also, from Problem 6a, $x \vee I = I$, i.e. $x \subset I$. Hence, together, $O \subset x \subset I$ for any $x \in B$.

19. If $x \subset y$ and $y \subset x$, then $x = y$.

PROOF:

Since, by Problem 12 and hypothesis, $x \vee y = y$ and $y \vee x = x$, and also, by B2a, $x \vee y = y \vee x$, it immediately follows that $x = y$.

20. $x \subset y$ iff $y' \subset x'$.

PROOF:

$$x \subset y \leftrightarrow x \wedge y' = O \leftrightarrow y' \wedge x = O \leftrightarrow y' \wedge x'' = O \leftrightarrow y' \subset x'$$

i.e. $x \subset y \leftrightarrow y' \subset x'$, by B6, 2a, Prob. 5, B6.

21. If $x \subset z$ and $y \subset z$, then $(x \vee y) \subset z$.

PROOF:

Since $x \vee z = z$ and $y \vee z = z$, by hypothesis and Problem 12, it follows that

$$(x \vee y) \vee z = x \vee (y \vee z) = x \vee z = z, \quad \text{i.e. } (x \vee y) \subset z$$

22. A Boolean lattice L is a Boolean algebra B .

PROOF:

The transformation is clear through the following 1-1 correspondence:

	L	B
1. Commutative law:	L1	\leftrightarrow B2
2. Associative law:	L2	\leftrightarrow Prob. 9
3. Absorption law:	L3	\leftrightarrow Prob. 4
4. Idempotent law:	L4	\leftrightarrow Prob. 4
5. Distributive law:	L5	\leftrightarrow B3
6. Complementation:	Df. 1.4.1.14	\leftrightarrow B5

23. B becomes L under **B6** (cf. Th. 2.4.2.4).

PROOF:

- (i) B is partly ordered under **B6**, since, for $x, y, z \in B$,
- (a) Reflexive: $x \subset x$, by Problem 13;
 - (b) Antisymmetric: $x \subset y$ and $y \subset x$ imply $x = y$, by Problem 19;
 - (c) Transitive: $x \subset y$ and $y \subset z$ imply $x \subset z$, by Problem 17.
- (ii) $x \vee y \in B$, by **B1**, is also an upper bound of x and y , since $x \vee (x \vee y) = (x \vee x) \vee y = x \vee y$, by Prob. 9, 4, which proves, by Prob. 12, $x \subset x \vee y$, and likewise $y \subset x \vee y$.
- (iii) Let b be any upper bound of x and y , i.e. $x \subset b$ and $y \subset b$; then, by Prob. 12, $x \vee b = b$ and $y \vee b = b$. Hence $(x \vee y) \vee b = x \vee (y \vee b) = x \vee b = b$, i.e., again by Prob. 12, $x \vee y \subset b$, which proves that $x \vee y$ is the l.u.b., i.e. the join, of x and y .

The steps of (ii) and (iii) can be similarly taken for $x \wedge y$, to prove it to be the g.l.b., i.e. the meet, of x and y .

Hence B under **B6** is a lattice.

*24. Find, by Df. 2.4.2.5, the diagrams of circuits which correspond to the following propositions:

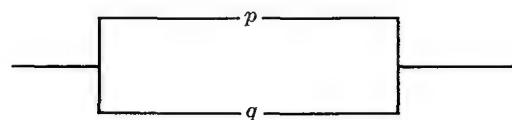
- (i) $p \vee q$, (ii) pq , (iii) $(p \vee (\overline{p \vee q})) \vee pq$.

Solution:

- (i) Since a switching circuit design is an arrangement of wires and switches where an *open* switch prevents the flow of current while a *closed* switch permits the flow, the table at right exhausts all possible cases, given two distinct switches p and q through which current is to flow if either p or q or both are closed. It is evident that the first table is logically equivalent to the second table, which is the truth-table of $p \vee q$ where p and q are two propositions. The circuit for $p \vee q$ is then represented by the two switches p and q in parallel as at right.

p	q	parallel circuit
on	on	on
on	off	on
off	on	on
off	off	off

p	q	$p \vee q$
1	1	1
1	0	1
0	1	1
0	0	0



- (ii) Likewise, there exists a 1-1 correspondence between the conjunction of two propositions pq and the condition that current flow if both switches p and q are closed, as is manifest in the corresponding two tables below:

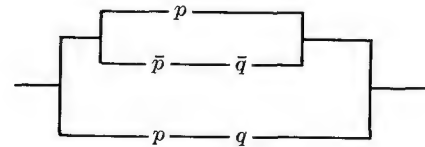
p	q	series circuit
on	on	on
on	off	off
off	on	off
off	off	off

p	q	pq
1	1	1
1	0	0
0	1	0
0	0	0

The circuit for pq is then represented by p and q in series, i.e. as follows:



- (iii) Once it is proved, by (i) and (ii), that there exists a 1-1 correspondence between a disjunction and a parallel circuit and between a conjunction and a series circuit, and, by Df. 2.4.2.5, that \bar{p} is on and off if p is off and on respectively, the diagram of the circuit is immediately obtained for $(p \vee (\bar{p} \vee q)) \vee pq$, or what is the same, $(p \vee (\bar{p} \wedge \bar{q})) \vee (p \wedge q)$, as shown in the adjoining diagram.

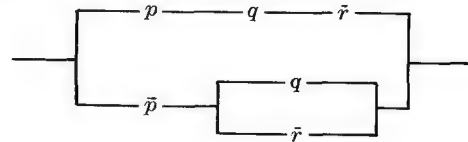


***25. Design circuits for the following propositions:**

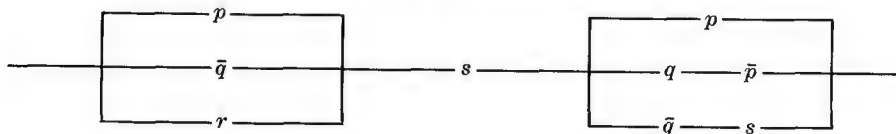
- (i) $(pq\bar{r}) \vee (\bar{p}(q \vee \bar{r}))$, (ii) $(p \vee \bar{q} \vee r) s (p \vee q\bar{r} \vee \bar{q}s)$.

Solution:

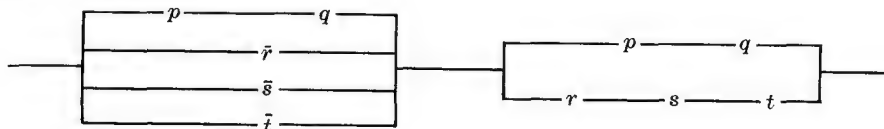
- (i) Since $pq\bar{r}$ is a series circuit, which is in parallel with $\bar{p}(q \vee \bar{r})$ which in turn is \bar{p} in series with a parallel circuit of $q \vee \bar{r}$, the proposition reveals itself in the adjoining figure.



- (ii) Reasoning similarly, the second design is obtained as follows:



***26. Represent the circuit below in a proposition, then simplify it by the theorems discovered in §1.1.1.**



Solution:

Since the first parallel circuits are represented by $(pq \vee \bar{r} \vee \bar{s} \vee \bar{t})$ and the second by $(pq \vee rst)$, and since they are in series, the design as a whole is represented by $(pq \vee \bar{r} \vee \bar{s} \vee \bar{t})(pq \vee rst)$. Now

$$\begin{aligned}
 (pq \vee \bar{r} \vee \bar{s} \vee \bar{t})(pq \vee rst) &\equiv ((pq) \vee (\bar{r} \vee \bar{s} \vee \bar{t}))(pq \vee (rst)) && \text{by Df.} \\
 &\equiv (pq) \vee ((\bar{r} \vee \bar{s} \vee \bar{t})(rst)) && \text{by Prob. 10, iii, a} \\
 &\equiv (pq) \vee ((\bar{rst})(rst)) && \text{by Prob. 12, ii} \\
 &\equiv pq && \text{by Prob. 14, ix of §1.1.1}
 \end{aligned}$$

Hence the given design is logically equivalent to the following design:

27. Prove Th. 2.4.2.7.

PROOF:

By Df. 2.4.2.6, B1-6 can be written as follows:

P1: Closure	(1a) $p \vee q \in P$ if $p, q \in P$	(1b) $pq \in P$ if $p, q \in P$
P2: Commutative law	(2a) $p \vee q \equiv q \vee p$	(2b) $pq \equiv qp$
P3: Distributive law	(3a) $p \vee (qr) \equiv (p \vee q)(p \vee r)$	(3b) $p(q \vee r) \equiv pq \vee pr$
P4: Identity	(4a) $p \vee q\bar{q} = p$	(4b) $pp \equiv p$
P5: Complement	(5a) $p \vee \bar{p}$ is a tautology.	(5b) $p\bar{p}$ is a contradiction.
P6: Inclusion	(6) $p \rightarrow q \equiv p \vee \bar{q}$	

Since it is already known that every proposition included in P1-6 is a tautology (which can be readily verified by truth-tables as in §1.1.1), this completes the proof.

Supplementary Problems

Part 2

2.1. Prove that the null set \emptyset (cf. Df. 2.1.7) is a subset of every set.2.2. Prove, first in terms of ε (membership), then by a Venn diagram, that

$$A \cap (B \cap C) = (A \cap B) \cap C = A \cap B \cap C$$

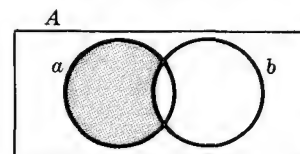
2.3. Give second proofs of Th. 2.3.7 in entirety (except the one already given) by Venn diagrams.

2.4. For any two sets A and B ,

$$(i) \ A \cup (B - A) = A \cup B, \quad (ii) \ A \cap (B - A) = \emptyset$$

2.5. If A and B are two subsets of a set C , and if $A \cup B = C$ and $A \cap B = \emptyset$, then $B = C - A$.2.6. Prove that $\{\{a\}, \{a, b\}\} = \{\{c\}, \{c, d\}\}$ iff $a = c$ and $b = d$.2.7. Given $X = \{p, q\}$ and $Y = \{r, s, t\}$, find

$$(i) \ X \times X, \quad (ii) \ Y \times Y, \quad (iii) \ X \times Y, \quad (iv) \ Y \times X.$$

2.8. Prove that the four categorical propositions — A (all a is b), E (no a is b), I (some a is b), and O (some a is not b) — can be expressed as $a\bar{b} = 0$, $ab = 0$, $ab \neq 0$, and $a\bar{b} \neq 0$ respectively.2.9. Consider Prob. 2.8 above in terms of Venn diagrams, interpreting them in the language of ε as in the figure at right.

$$(Ex)((x \varepsilon a)(x \varepsilon \bar{b}))$$

2.10. Simplify the following expressions, justifying each step with B1-6 and other theorems deduced from them:

$$(i) \ p \vee (p \wedge (p \vee r)) \vee q \wedge (p \vee q').$$

$$(ii) \ ((p \wedge q') \vee (p' \wedge q))' \wedge ((p \wedge q) \vee (p' \wedge q'))'.$$

$$(iii) \ ((p \wedge q) \vee (p \wedge r) \vee (p' \wedge s' \wedge t')) \wedge ((p \wedge q' \wedge r) \vee (p' \wedge s' \wedge t') \vee (p' \wedge q' \wedge t'))'.$$

2.11. Verify the validity of the results of Prob. 2.10 by truth-tables or (in particular for (iii)) by meta-theorems of §1.1.1.

2.12. Simplify, justify, and verify (as in Prob. 2.10-11) the following expressions:

- (i) $(p \wedge q' \wedge r')' \wedge (p' \wedge q \wedge r)'$.
- (ii) $(p \vee q \vee r \vee (p' \wedge q' \wedge r')) \wedge ((p \wedge q) \vee (p \wedge q') \vee (p' \wedge q))$.
- (iii) $((p \wedge q) \vee (r \wedge s \wedge t)) \wedge (r' \vee s' \vee t' \vee (p \wedge q))$.

2.13. If $o(X)$ denotes the number of elements in a set X (cf. Df. 2.1.11), and if A and B are any two sets, then

$$o(A \cup B) = o(A) + o(B) - o(A \cap B)$$

2.14. If A_1, A_2, \dots, A_n are mutually disjoint sets such that $A_1 \cup A_2 \cup \dots \cup A_n = U$, (cf. Df. 2.3.13), then for any set B ,

$$o(B) = o(A_1 \cap B) + o(A_2 \cap B) + \dots + o(A_n \cap B)$$

2.15. Draw a Venn diagram for the case of three subsets A, B, C , and define the eight disjoint (nonoverlapping) regions R_1, R_2, \dots, R_8 in terms of " ε " (e.g. $R_1 = (x \varepsilon A) \wedge (x \varepsilon B) \wedge (x \varepsilon C)$, \dots , $R_8 = (x \notin A) \wedge (x \notin B) \wedge (x \notin C)$), then determine the following sets by listing their elements in terms of R_i , $i = 1, 2, \dots, 8$:

- (i) U , (ii) A , (iii) B , (iv) C , (v) A' , (vi) $A \cup B$, (vii) $A \cap B$
- (viii) $A' \cap (A \cap B)$, (ix) $(A \cup B) \cap C$, (x) $(A \cap B') \cap C'$

2.16. A foreign language school has 200 students, of which 120 students study French, denoted by $o(F) = 120$ (cf. Prob. 2.13), 90 students study German, i.e. $o(G) = 90$, and 70 students study Russian, i.e. $o(R) = 70$. It is also known that $o(R \cap G) = 30$ (i.e. 30 students study both Russian and German), $o(R \cap F) = 50$, $o(G \cap F) = 40$, $o(F \cap G \cap R) = 20$. Find $o(R \cap G' \cap F')$ and $o(R' \cap G \cap F)$.

2.17. The result of a poll shows that the numbers of people who listen to the programs A , B , and C are a , b , and c respectively; and the numbers of people who listen to both A and B , both B and C , and both C and A are d , e , and f respectively. Find the number of people who listen to A , B , and C .

2.18. Draw the switching circuits which represent the following expressions:

- (i) $p \wedge (q \wedge (r \vee s) \vee (r \wedge (t \vee u)))$, (ii) $(p \wedge q \wedge r) \vee (p \wedge q \wedge ((r \wedge s) \vee (t \wedge u)))$.

*2.19. Referring to Prob. 2.13 above, interpret the following set of axioms

- (i) $P(x) = 0$ if x is a logically false proposition.
 - (ii) $0 \leq P(x) \leq 1$ for any proposition x .
 - (iii) $P(x \vee y) = P(x) + P(y) - P(xy)$ for any two propositions x and y ,
- where $P(x)$ denotes the *probability* of any proposition x and the same probability is assigned to any two equivalent propositions.

*2.20. If x and y in Prob. 2.19, (iii) are inconsistent, then

$$(iv) \quad P(x \vee y) = P(x) + P(y)$$

and also, \bar{x} designating the negated x ,

$$(v) \quad P(\bar{x}) = 1 - P(x)$$

Finite Groups

§3.1.1 Groups in General

Df. 3.1.1.1 A *group* is a set G of elements a, b, c, \dots under a binary operation $*$ (or \circ or any other suitable symbol or, as below, no visible symbol at all), satisfying the following four axioms:

- G1:** Closure. $a * b \in G$ (or more simply written: $ab \in G$) is unique for every $a, b \in G$; $a * b$ or ab is called the *product* of the factors a and b .
- G2:** Associativity. $a(bc) = (ab)c$ for every $a, b, c \in G$.
- G3:** Identity. $e \in G$ such that $ea = ae = a$ for every $a \in G$.
- G4:** Inverse. $a^{-1} \in G$ such that $a^{-1}a = aa^{-1} = e$ for every $a \in G$.

Th. 3.1.1.2 The identity of **G3** and the inverse of **G4**, defined by Df. 3.1.1.1, are unique. (Cf. Prob. 1.)

Th. 3.1.1.3 (Cancellation Law for Groups). For every $a, b, x \in G$, both $xa = xb$ and $ax = bx$ imply the same $a = b$. (Cf. Prob. 2.)

The definition of groups by Df. 3.1.1.1 is in fact *stronger* than necessary, since it contains some redundant properties, viz. one-half of **G3** and **G4**. These as such can be replaced by the corresponding *weaker* axioms as in Th. 3.1.1.4 below.

Th. 3.1.1.4 Df. 3.1.1.1 is equivalent to the following alternative set of axioms:

- G1'** \equiv **G1**
- G2'** \equiv **G2**
- G3'** Left-identity: $ea = a$, $e \in G$, for every $a \in G$.
- G4'** Left-inverse: $a^{-1}a = e$, $a^{-1} \in G$, for every $a \in G$.

The term “left” here is the dual of “right”; for the left-identity can be replaced by a right-identity and the left-inverse by a right-inverse (cf. Prob. 3).

Df. 3.1.1.1 can be made even weaker, abstracting away the axiom of the explicit identity, as in Th. 3.1.1.5 below.

Th. 3.1.1.5 Df. 3.1.1.1 is equivalent to the following alternative set of axioms (cf. Prob. 4):

- G1''** \equiv **G1**
- G2''** \equiv **G2**
- G3''** Unary operation “inverse”: $a^{-1} \in G$ is unique for every $a \in G$.
- G4''** Inverse law: $a^{-1}(ab) = b = (ba)a^{-1}$.

The number of axioms may be reduced, too, and the following set of axioms for G is an example.

Th. 3.1.1.6 Df. 3.1.1.1 is equivalent to:

$$\mathbf{G1}''' \equiv \mathbf{G1}$$

$$\mathbf{G2}''' \equiv \mathbf{G2}$$

$$\mathbf{G3}''' \equiv \text{Th. 3.1.1.3 for both left and right; viz. each of } xa = xb \text{ and } ay = by \text{ for every } a, b, x, y \in G \text{ implies } a = b. \text{ (Cf. Prob. 5, 11.)}$$

Not only other alternative sets of axioms for G are available (cf. Prob. 6 and Supplementary Prob. 3.3), but also, depending on the number of axioms involved, other kinds of groups may be obtained.

Df. 3.1.1.7 A group G , defined by $\mathbf{G1-4}$ (or other equivalent alternative sets of axioms), is called an *Abelian* (or a *commutative*) group if it satisfies an additional axiom:

G5. Commutative law: $ab = ba$ for every $a, b \in G$.

In particular, an Abelian group under addition is sometimes called a *module*.

Note that an Abelian group is a special group; if a set G is an Abelian group, it is then *a fortiori* a group, satisfying $\mathbf{G1-4}$. In this sense the class \mathbf{A} of all Abelian groups is a sub-class of the class \mathbf{G} of all the sets which satisfy $\mathbf{G1-4}$.

On the other hand, the class \mathbf{G} of ordinary groups may become a sub-class of the class of more general groups such as semi-groups, quasi-groups, loops, etc., defined as below.

Df. 3.1.1.8 If a set D satisfies only $\mathbf{G1}$ and $\mathbf{G2}$, it is then called a *semi-group* (or *demi-group* in the French mathematical literature in which a semi-group satisfies $\mathbf{G1-2}$ and Th. 3.1.1.3). (Cf. Prob. 11.)

In terms of “groupoid” (cf. Df. 2.2.1.1b), the semi-group may be defined also as follows:

Df. 3.1.1.8a A *semi-group* is an associative groupoid (cf. Df. 2.2.1.4).

It must be noted that a semi-group which also satisfies $\mathbf{G3}$ is called a *monoid*; viz. a monoid is an associative groupoid with an identity element.

Df. 3.1.1.9 If a set Q satisfies $\mathbf{G1}''$ uniquely, i.e. if any two of $a, b, c \in Q$ in $ab = c$ uniquely determine the third, it is then called a *quasi-group*.

Df. 3.1.1.10 If a set L is a quasi-group with $\mathbf{G3}$, i.e. a two-sided identity: $ea = ae = a$ for any $a \in L$, it is then called a *loop*.

It is evident that G , defined by all of $\mathbf{G1-4}$, is a special kind of D, Q, L , satisfying each of Df. 3.1.1.8, Df. 3.1.1.9, and Df. 3.1.1.10.

Df. 3.1.1.11 A non-empty subset S of G under $*$ (a binary operation in Df. 3.1.1.1) is called a *complex* of G ; S is then called a *subgroup* of G if S is itself a group under $*$.

In particular, G is considered a subgroup of itself, and the unique set which consists of the identity e alone, which does form a group, as can be readily verified (cf. Prob. 15 below), is regarded as a subgroup of every group, including itself.

Df. 3.1.1.12 A subgroup which is neither G nor e alone is called a *proper* subgroup.

Th. 3.1.1.13 Any complex S of a group G is a subgroup iff

(i) $a, b \in S$ implies $ab \in S$, and

(ii) $a \in S$ implies $a^{-1} \in S$. (Cf. Prob. 8 and also Th. 3.2.1.1.)

Df. 3.1.1.14 If G has n elements, $n \in N$, N denoting as usual the set of all natural numbers, it is then called a *finite group of order n* .

Df. 3.1.1.15 If G has infinitely many elements, it is called a group of *infinite order* or simply an *infinite group*.

Example:

The set of all integers under addition; or the set of all rational numbers, excluding 0, under multiplication (cf. Prob. 9).

The groups in this chapter will be of finite order unless stated otherwise.

Solved Problems

1. The identity of **G3** and the inverse of **G4** are unique.

PROOF:

- (i) Suppose e' is also an identity; then $ee' = e'$, and since e is also an identity, $ee' = e$. Hence $ee' = e' = e$, which explicitly reveals the uniqueness of e .
- (ii) Suppose b is also an inverse of a , i.e. $ab = aa^{-1} = e$. Then, since $a^{-1} \in G$, it follows that $a^{-1}(ab) = a^{-1}(aa^{-1})$, by **G1**, and $(a^{-1}a)b = (a^{-1}a)a^{-1}$, by **G2**. Hence, by **G4** itself, $eb = ea^{-1}$, and $b = a^{-1}$, by **G3**, proving that the inverse a^{-1} is unique.

2. Prove the cancellation law for G (cf. Th.3.1.1.3).

PROOF:

Since, by **G1**, $xa = xb$ in G implies $x^{-1}(xa) = x^{-1}(xb)$, where $x^{-1} \in G$, it follows from **G2** that $(x^{-1}x)a = (x^{-1}x)b$, i.e. $ea = eb$, by **G4**. Hence, by **G3**, $a = b$; i.e. $xa = xb$ does imply $a = b$ in G .

Likewise $ax = bx$ in G implies $a = b$, which completes the proof.

3. Prove Th.3.1.1.4.

PROOF:

In accordance with the weakened axioms, the cancellation law also weakens; it becomes a left-cancellation law that $xa = xb$ implies $a = b$ (proved as above, in Prob. 2).

Now, the left-identity of **G3'** is also a right-identity, since, by **G4'**, $a^{-1}a = e = ee = a^{-1}ae$, i.e. $a^{-1}ae = a^{-1}a$ and, by left-cancellation, $ae = a$. Hence **G3'**, implicitly representing the two-sided identity, is equivalent to **G3**.

The left-inverse of **G4'** is also a right-inverse, since, by **G2', 3', 4'**, $a^{-1} = ea^{-1} = (a^{-1}a)a^{-1} = a^{-1}(aa^{-1}) = a^{-1}e$, i.e. $a^{-1}(aa^{-1}) = a^{-1}e$ and, by left-cancellation, $aa^{-1} = e$.

Hence, likewise, **G4'** is equivalent to **G4**.

Hence, altogether, the set **G1'-4'** is equivalent to **G1-4**.

Second Proof (without resorting to cancellation). Letting, by **G4'**, $a^{-1}a = e$ and $xa^{-1} = e$, then applying **G2', 3', 4'** repeatedly,

$$aa^{-1} = e(aa^{-1}) = (xa^{-1})(aa^{-1}) = x(a^{-1}(a(a^{-1}))) = x((a^{-1}a)a^{-1}) = x(ea^{-1}) = xa^{-1} = e$$

which proves the right-inverse for **G4'**. Using this result immediately,

$$ae = a(a^{-1}a) = (aa^{-1})a = ea = a$$

which proves the right-identity for **G3'**.

Note. It goes without saying that, conversely, **G3-4** entails **G3'-4'**, since the former explicitly contains the latter in itself.

Note also that another equivalent set of axioms for G can be readily obtained from **G1'-4'**, simply by replacing "left" by "right"; the proof will be carried out with (or without) right-cancellation.

4. Prove Th. 3.1.1.5.

PROOF:(i) As above, **G1-4** obviously entails **G1''-4''**.(ii) Conversely, the latter entails the former, since **G3-4** can be obtained from **G3''-4''** as follows. Substitute $c^{-1}c$ for b in $b = (ba)a^{-1}$ of **G4''**, and $c^{-1}c = ((c^{-1}c)a)a^{-1} = (c^{-1}c)(aa^{-1}) = c^{-1}(c(aa^{-1})) = aa^{-1}$. If $a = c$, then $aa^{-1} = a^{-1}a = x$ is unique in G for every $a = c \in G$; hence, if x as such is defined as the identity, **G4** is at once obtained.Furthermore, if $x = e$, then $xb = (a^{-1}a)b = a^{-1}(ab) = b$, by **G2''** and **G4''**; likewise $bx = b$, which yields **G3**.Hence the set **G1''-4''** is equivalent to the set **G1-4**.

5. Prove Th. 3.1.1.6.

PROOF:

Let

$$a_1, a_2, \dots, a_n \quad (A)$$

be the distinct elements of the group G of order n ; then, for any $a_i \in G$,

$$a_i a_1, a_i a_2, \dots, a_i a_n \quad (B)$$

are, by **G1'''**, also the elements of G and are distinct, since, by **G3'''**,

$$a_i a_j = a_i a_k \quad \text{implies} \quad a_j = a_k$$

Hence (B) is actually a different arrangement of (A), and if $a_x \in G$, there exists an element a_y such that

$$a_x = a_i a_y$$

Likewise, by **G3'''**, there exists a_z such that

$$a_x = a_z a_i$$

The identity element of G is then defined by letting $a_i = a_x = e$, since, given any two elements a_u and a_v of G such that

$$a_u a_i = a_i \quad \text{and} \quad a_i a_v = a_i$$

it always follows that

$$a_u a_x = a_u (a_i a_y) = (a_u a_i) a_y = a_i a_y = a_x$$

and likewise, $a_x a_v = a_x$. This establishes **G3**.Now, if there exist any two elements a_p and a_q in G such that

$$a_i a_p = e \quad \text{and} \quad a_q a_i = e,$$

then

$$a_q a_i a_p = (a_q a_i) a_p = e a_p = a_p$$

and

$$a_q a_i a_p = a_q (a_i a_p) = a_q e = a_q$$

Hence $a_p = a_q = a_i^{-1}$, where a_i^{-1} must be unique, since, by **G3'''**, there can be one and only one solution for $a_i a_x = e$. This yields **G4**, completing the proof.6. Prove that Df. 3.1.1.1 is equivalent to the following definition of a group: A group is a set G of elements a, b, c, \dots under a binary operation $/$, satisfying the five axioms below:**G'1.** $a/b \in G$ is unique for every $a, b \in G$.

$$\mathbf{G'4.} \quad (a/a)/(b/c) = c/b$$

G'2. $a/a = b/b = 1$

$$\mathbf{G'5.} \quad (a/b)/(c/b) = a/c$$

G'3. $a/(a/a) = a$ **PROOF:**(i) Referring to **G'2-3**, let $(a/a)/a = 1/a \equiv a^{-1}$; then, by **G'3-4**,

$$(a^{-1})^{-1} = (a^{-1}/a^{-1})/a^{-1} = (a^{-1}/a^{-1})/((a/a)/a) = a/(a/a) = a \quad \text{and} \quad ab = a(b^{-1})^{-1} = a/b^{-1}$$

Hence, since $a/b^{-1} \in G$ is unique, ab is unique, deducing **G1**.(ii) Since $a/b = a/(b^{-1})^{-1} = ab^{-1}$, $a/a = aa^{-1} = 1$; and, immediately using this result, $a^{-1}a = a^{-1}(a^{-1})^{-1} = 1$. Hence $aa^{-1} = a^{-1}a = 1$, establishing **G4**.

- (iii) Let $a = 1$ in $a^{-1} = (a/a)/a$; then $1^{-1} = 11^{-1}$, and $1 = 1/1 = 11^{-1} = 1^{-1}$. Hence, by G'2, $a1^{-1} = a1 = a$ and, by definition, $a^{-1} = 1/a = 1a^{-1}$, from which it follows that $(a^{-1})^{-1} = 1(a^{-1})^{-1}$, i.e. $a = 1a$, proving G3.
- (iv) Let $a \equiv x$, $b \equiv y^{-1}$, $c \equiv 1$ in G'5; then $(xy)(1y)^{-1} = x1^{-1}$, i.e. $(xy)y^{-1} = x$. Now, let $a \equiv xy$, $b \equiv y$, $c \equiv z^{-1}$; then $ab^{-1} = (xy)y^{-1} = x$ and G'5 itself becomes $(ab^{-1})(cb^{-1})^{-1} = ac^{-1}$. Hence, since $(bc^{-1})^{-1} = cb^{-1}$ (G'4: $1(bc^{-1})^{-1} = cb^{-1}$), it follows that $(ab^{-1})(bc^{-1}) = ac^{-1}$. This, in terms of x, y, z , is indeed $x(yz) = (xy)z$, proving G2.

Conversely, through $ab^{-1} = a/b$, G'1-5 can be deduced from G1-4, establishing the desired equivalence between the original and the alternative definitions.

7. If $aa = a$, $a \in G$, then $a = e$.

PROOF:

$$a = ae = a(aa^{-1}) = (aa)a^{-1} = aa^{-1} = e$$

8. Prove Th. 3.1.1.13.

PROOF:

If (i) and (ii) hold for S , then S immediately satisfies G1, by (i) itself, and (i) also assures G2 for S . For, if $a, b \in S$ implies $ab \in S$, then both $ab, c \in S$ and $a, bc \in S$ imply the same: $abc \in S$. Also, since there is at least one element in S , $e = aa^{-1}$ must be in S , by (ii), proving G3 for S . As for G4, it is directly provided by (ii) itself and the existence of e , which has already been established.

Conversely, if S is a subgroup of G , (i) obviously holds. Also, since the identity $x = e'$ of S satisfies $xx = x$ (cf. Prob. 7), it is the identity of G itself. But then, since the inverse of any $a \in G$ is unique, the inverse of any element $s \in S$ must be the same as its inverse in G . Hence (ii) holds.

This completes the proof.

9. Prove that the set I of all integers under addition is a group, and so is the set R of all rational numbers ($0 \notin R$) under multiplication.

PROOF:

- (i) G1: $x, y \in I$ implies $x + y \in I$. G2: $x, y, z \in I$ implies $x + (y + z) = (x + y) + z$. G3: $e = 0$. G4: $x^{-1} = -x$.

- (ii) G1: $x, y \in R$ implies $xy \in R$. G2: $x, y, z \in R$ implies $x(yz) = (xy)z$. G3: $e = 1$. G4: $x^{-1} = 1/x$.

Note that both I and R are Abelian groups, since $(x + y) = (y + x) \in I$ and $(xy) = (yx) \in R$, satisfying G5.

10. Why, or why not, are the following sets groups under multiplication?

- (i) The set I of all integers.
 (ii) The set C of all complex numbers.
 (iii) The set S of all real numbers of the form $x + y\sqrt{2}$, where $x, y \in R$ are not simultaneously zero.
 (iv) The set U of the third roots of unity.

PROOF:

- (i) I is not a group under multiplication, because G4 does not hold here.

- (ii) C forms a group under multiplication, since

G1: $(a+bi), (c+di) \in C$, where $a, b, c, d \in R^*$ (R^* denoting as before the set of all real numbers) implies $(a+bi)(c+di) = (ac-bd) + (ad+bc)i \in C$, where $(ac-bd), (ad+bc) \in R^*$.

G2: $(a+bi), (c+di), (e+fi) \in C$, where $a, b, c, d, e, f \in R^*$, implies

$$\begin{aligned} (a+bi)((c+di)(e+fi)) &= (ace - adf - bcf - bde) + (acf + ade + bce - bdf)i \\ &= ((a+bi)(c+di))(e+fi) \end{aligned}$$

G3: $e = 1 (= 1 + 0i)$.

G4: $(a+bi)^{-1} = 1/(a+bi)$.

(iii) S is a group under multiplication, since

G1: $(a + b\sqrt{2}), (c + d\sqrt{2}) \in S$, where $a, b, c, d \in R$, implies

$$(a + b\sqrt{2})(c + d\sqrt{2}) = (ac + 2bd) + (ad + bc)\sqrt{2} \in S, \text{ where } (ac + 2bd), (ad + bc) \in R.$$

G2: $(a + b\sqrt{2}), (c + d\sqrt{2}), (e + f\sqrt{2}) \in S$, where $a, b, c, d, e, f \in R$, implies

$$\begin{aligned} (a + b\sqrt{2})((c + d\sqrt{2})(e + f\sqrt{2})) &= (ace + 2adf + 2bce + 2bde) + (acf + ade + bce + 2bdf)\sqrt{2} \\ &= ((a + b\sqrt{2})(c + d\sqrt{2}))(e + f\sqrt{2}) \end{aligned}$$

G3: $e = 1 (= 1 + 0\sqrt{2})$.

G4: $(a + b\sqrt{2})^{-1} = 1/(a + \sqrt{2}) = (a/(a^2 - 2b^2)) + (-b/(a^2 - 2b^2))\sqrt{2}$, $a/(a^2 - 2b^2), -b/(a^2 - 2b^2) \in R$ (since $a^2 \neq 2b^2$, because $a^2 = 2b^2$ will mean $a = \pm b\sqrt{2} \notin R$, contrary to the initial condition).

(iv) Solve $x^3 - 1 = 0$, i.e. $(x - 1)(x^2 + x + 1) = 0$, and the third roots of unity are: $1, (-1 + i\sqrt{3})/2, (-1 - i\sqrt{3})/2$. Let $a \equiv (-1 + i\sqrt{3})/2$ and $b \equiv (-1 - i\sqrt{3})/2$; then, since $ab = ba = 1$, it follows that **G1-2**, **G3**, and **G4** ($1^{-1} = 1, a^{-1} = b, b^{-1} = a$) are all satisfied. Hence the set U forms a group.

11. A semi-group D of order n is a group if, for $a, x, y \in D$,

(i) $ax = ay$ implies $x = y$, (ii) $xa = ya$ implies $x = y$

PROOF:

Let the n elements of D be d_1, d_2, \dots, d_n ; then, by (i), ad_1, ad_2, \dots, ad_n are all distinct elements of D (cf. Prob. 5). Hence, for $a, b \in D$, there uniquely exists d_i such that $ad_i = b$, which implies that $ay = b$ has a unique solution d_i .

Likewise, by (ii), $xa = b$ also has a unique solution.

Hence **G3'''** is established, and since **G1'''-2'''** are already given by definition, D is a group by Th. 3.1.1.6.

12. Given $axa = b$ in G , find x .

Solution:

Multiplying both sides of the equation on the left, then on the right, by a^{-1} ,

$$a^{-1}(axa) = a^{-1}b \rightarrow xa = a^{-1}b \rightarrow xaa^{-1} = a^{-1}ba^{-1} \rightarrow x = a^{-1}ba^{-1}$$

13. Prove: $(a^{-1})^{-1} = a$, where $a \in G$.

PROOF:

Since, by **G4**, $(a^{-1})^{-1}a^{-1} = e$, multiply both sides of the equation on the right by a to obtain $((a^{-1})^{-1}a^{-1})a = ea = a$. But, by **G2-4**,

$$((a^{-1})^{-1}a^{-1})a = (a^{-1})^{-1}(a^{-1}a) = (a^{-1})^{-1}e = (a^{-1})^{-1}$$

Hence $(a^{-1})^{-1} = a$.

14. Define $a^0 = e$, where $a \in G$; then prove that for any integer n , $a^{-n} = (a^{-1})^n = (a^n)^{-1}$.

PROOF:

(i) $n > 0$. For $n = 1$, $a^{-1} = (a^{-1})^1 = (a^1)^{-1}$, which obviously holds.

Suppose $a^{-k} = (a^{-1})^k = (a^k)^{-1}$ for $n = k$; then, for $n = k + 1$,

$$a^{-(k+1)} = a^{-k}a^{-1} = (a^{-1})^k(a^{-1}) = (a^{-1})^{k+1}$$

and since $a^{-(k+1)} = (a^{-1})^k(a^{-1}) = (a^k)^{-1}(a^{-1})$, it follows that

$$a^{k+1}a^{-(k+1)} = aa^k(a^k)^{-1}(a^{-1}) = aea^{-1}aa^{-1} = e, \text{ i.e. } a^{-(k+1)} = (a^{k+1})^{-1}$$

Hence $a^{-(k+1)} = (a^{-1})^{k+1} = (a^{k+1})^{-1}$. Thus in general, $a^{-n} = (a^{-1})^n = (a^n)^{-1}$ for any integer $n > 0$.

(ii) $n < 0$. Let $n = -r$ ($r > 0$); then $a^{-n} = a^r$, and $(a^{-1})^n = (a^{-1})^{-r} = ((a^{-1})^{-1})^r = a^r$ (cf. Prob. 13); $(a^n)^{-1} = (a^{-r})^{-1} = ((a^r)^{-1})^{-1} = a^r$. Hence, again, $a^{-n} = (a^{-1})^n = (a^n)^{-1}$ for any integer $n < 0$.

(iii) For $n = 0$, $a^{-0} = a^0 = e$; $(a^{-1})^0 = e$ (since $a^0 = e$ and $aa^{-1} = e$); $(a^0)^{-1} = e^{-1} = e$. Hence $a^{-0} = (a^{-1})^0 = (a^0)^{-1}$.

Hence, for any integer n , $a^{-n} = (a^{-1})^n = (a^n)^{-1}$.

15. Prove: $e^n = e$, $e \in G$, where $e^0 = e$, for any integer n .

PROOF:

(i) $n > 0$. For $n = 1$, $e^1 = e$. Suppose $e^k = e$ for $n = k$; then, for $n = k + 1$, $e^{k+1} = e^k e = ee = e$.
Hence $e^n = e$ for any integer $n > 0$.

(ii) $n < 0$. Since $ee = e$, $e^{-1} = e$ for $n = -1$. In general, letting $n = -r$ ($r > 0$), $e^n = e^{-r} = (e^{-1})^r = e^r = e$.

Hence $e^n = e$ for any integer n .

16. Generalize G2, i.e. $(a_1 a_2 \dots a_m)(a_{m+1} a_{m+2} \dots a_{m+n}) = a_1 a_2 \dots a_{m+n}$.

PROOF:

Since, by G2,

$$a_1(a_2 a_3) = (a_1 a_2)a_3 = a_1 a_2 a_3, \quad (a_1(a_2 a_3))a_4 = a_1 a_2 a_3 a_4, \quad \dots, \quad (a_1(a_2(\dots a_{n-1})))\dots a_n = a_1 a_2 \dots a_n$$

let $\prod_{i=1}^1 a_i = a_1$ and $\prod_{i=1}^{m+1} a_i = \left(\prod_{i=1}^m a_i\right) a_{m+1}$. Then, since the case for $n = 1$ evidently holds, suppose

$$\prod_{i=1}^m a_i \prod_{k=1}^n a_{m+k} = \prod_{i=1}^{m+n} a_i \quad \text{for } n.$$

Then, for $n + 1$,

$$\begin{aligned} \prod_{i=1}^m a_i \prod_{k=1}^{n+1} a_{m+k} &= \prod_{i=1}^m a_i \left(\left(\prod_{k=1}^n a_{m+k} \right) a_{m+n+1} \right) = \left(\prod_{i=1}^m a_i \prod_{k=1}^n a_{m+k} \right) a_{m+n+1} \\ &= \left(\prod_{i=1}^{m+n} a_i \right) a_{m+n+1} = \prod_{i=1}^{m+n+1} a_i \end{aligned}$$

completing the proof.

17. For any integers m and n , and $a \in G$, $a^{m+n} = a^m a^n$.

PROOF:

(i) $m > 0, n > 0$: $a^{m+n} = \underbrace{a \dots a}_{m+n} = \underbrace{a \dots a}_m \underbrace{a \dots a}_n = a^m a^n$.

(ii) $m < 0, n < 0$: Let $m = -p$ and $n = -q$, $p > 0, q > 0$; then, by Prob. 14,
 $a^{m+n} = a^{-(p+q)} = (a^{-1})^{p+q} = (a^{-1})^p (a^{-1})^q = a^{-p} a^{-q} = a^m a^n$

(iii) $m = 0, n \neq 0$: $a^{0+n} = a^n = ea = a^0 a^n$. (Likewise when $m \neq 0, n = 0$.)

(iv) $m > 0, n < 0, m+n \geq 0$: Since $-n > 0$, it follows by (i) and (iii) of Prob. 14,

$$a^{m+n} = a^{m+n} e = a^{m+n} a^{-n} a^n = (a^{m+n} a^{-n}) a^n = a^{(m+n)+(-n)} a^n = a^m a^n$$

(v) $m > 0, n < 0, m+n < 0$: Since $-m-n > 0$, it follows from (i) that $a^{-m-n} a^m = a^{-m-n+m} = a^{-n}$, then, by Prob. 14, $a^n a^{-n} = a^{m+n} a^{-m-n} = e$ and

$$a^{m+n} = a^{m+n} e = a^{m+n} a^{-n} a^n = a^{m+n} (a^{-m-n} a^m) a^n = (a^{m+n} a^{-m-n}) a^m a^n = a^m a^n$$

(vi) The same result is obtained if m and n are interchanged in (iv) and (v).

This exhausts all possible cases, completing the proof.

18. For any integers m and n , and every $a \in G$, $(a^m)^n = a^{mn}$.

PROOF:

(i) $n > 0$: For $n = 1$, $(a^m)^1 = a^m = a^{m \cdot 1}$. If $(a^m)^k = a^{mk}$ for $n = k$, then, for $n = k + 1$, $(a^m)^{k+1} = (a^m)^k a^m = a^{mk} a^m = a^{mk+m} = a^{m(k+1)}$, proving the general case for $n > 0$.

(ii) $n < 0$: Let $n = -r$; then $r > 0$ and, by Prob. 14, $(a^m)^n = (a^m)^{-r} = ((a^m)^r)^{-1} = (a^{mr})^{-1} = a^{-mr} = a^{mn}$.

(iii) $n = 0$: $(a^m)^0 = e = a^0 = a^{m \cdot 0}$. This completes the proof.

19. In G , $(ab)^{-1} = b^{-1} a^{-1}$, and in general: $(a_1 a_2 \dots a_n)^{-1} = a_n^{-1} \dots a_2^{-1} a_1^{-1}$.

PROOF:

(i) Since $(ab)(ab)^{-1} = e$ and $(ab)b^{-1}a^{-1} = a(bb^{-1})a^{-1} = aea^{-1} = aa^{-1} = e$, i.e. $(ab)(ab)^{-1} = (ab)b^{-1}a^{-1}$, it follows, by right-cancellation, that $(ab)^{-1} = b^{-1}a^{-1}$.

(ii) When $n = 1$, $(a_1)^{-1} = a_1^{-1}$, which evidently holds. Suppose for $n = k$, $(a_1 a_2 \dots a_k)^{-1} = a_k^{-1} \dots a_2^{-1} a_1^{-1}$, i.e. $(a_1 a_2 \dots a_k)(a_k^{-1} \dots a_2^{-1} a_1^{-1}) = e$. Then, for $n = k + 1$,

$$\begin{aligned} (a_1 a_2 \dots a_k a_{k+1})(a_{k+1}^{-1} a_k^{-1} \dots a_2^{-1} a_1^{-1}) &= (a_1 a_2 \dots a_k) a_{k+1} a_{k+1}^{-1} (a_k^{-1} \dots a_2^{-1} a_1^{-1}) \\ &= (a_1 a_2 \dots a_k) e (a_k^{-1} \dots a_2^{-1} a_1^{-1}) \\ &= \dots = a_1 a_1^{-1} = e \end{aligned}$$

completing the generalization.

20. If $a, b \in G$, then $(bab^{-1})^n = ba^n b^{-1}$ for any integer n .

PROOF:

- (i) $n > 0$: If $n = 1$, it obviously holds. Suppose $(bab^{-1})^k = ba^k b^{-1}$; then, for $n = k + 1$, $(bab^{-1})^{k+1} = (bab^{-1})^k bab^{-1} = ba^k b^{-1} bab^{-1} = ba^k ab^{-1} = ba^{k+1} b^{-1}$. Hence $(bab^{-1})^n = ba^n b^{-1}$.
- (ii) $n < 0$: If $n = -1$, then $(bab^{-1})^{-1} = ba^{-1} b^{-1}$ since $bab^{-1} ba^{-1} b^{-1} = e$. Hence
- $$(bab^{-1})^n = (ba^{-1} b^{-1})^{-n} = b(a^{-1})^{-n} b^{-1} = ba^n b^{-1}$$

21. A group G is Abelian if, for $a, b \in G$, (i) $a^2 = e$, or (ii) $(ab)^2 = a^2 b^2$, or (iii) $b^{-1} a^{-1} ba = e$.

PROOF:

- (i) Since $a^2 = aa = e$ and $aa^{-1} = e$, it follows that $aa = aa^{-1}$ and, by left-cancellation, $a = a^{-1}$. Hence $(ab) = (ab)^{-1}$ and, by Prob. 19, $(ab)^{-1} = b^{-1} a^{-1} = ba$, i.e. $ab = ba$ ($= (ab)^{-1}$), establishing the commutativity in G .
- (ii) Writing it out, $(ab)^2 = (ab)(ab) = aabb = a^2 b^2$, i.e. $abab = aabb$. Then, by left-cancellation, $bab = abb$ and, by right-cancellation, $ba = ab$, which is G5 for an Abelian group.
- (iii) $ab = abe = ab(b^{-1} a^{-1} ba) = a(bb^{-1})a^{-1} ba = (aa^{-1})ba = ba$.

22. If a group G is Abelian and $a, b \in G$, then $(ab)^n = a^n b^n$ for any integer n .

PROOF:

- (i) $n > 0$: If $n = 1$, then $(ab)^1 = ab = a^1 b^1$, which is evidently true. Suppose, for $n = k$, $(ab)^k = a^k b^k$; then, for $n = k + 1$, $(ab)^{k+1} = (ab)^k ab = a^k b^k ab = a^k ab^k b = a^{k+1} b^{k+1}$, justifying the generalization.
- (ii) $n < 0$: Since $-n > 0$, it follows from (i) that $(ab)^{-n} = a^{-n} b^{-n}$. Also, by Problem 17, $(ab)^n (ab)^{-n} = e$ and $a^n b^n (ab)^{-n} = a^n b^n a^{-n} b^{-n} = a^n a^{-n} b^n b^{-n} = ee = e$, i.e. $(ab)^n (ab)^{-n} = a^n b^n (ab)^{-n}$. Hence, by right-cancellation, $(ab)^n = a^n b^n$.
- (iii) $n = 0$: $(ab)^0 = e = ee = a^0 b^0$, completing the generalization.

23. If $M(x) = rx + s$, where $x \in R^*$ (the set of all real numbers), $r, s \in R$ (the set of all rational numbers), and $r \neq 0$, then the set S of the function of the form M is a non-Abelian group.

PROOF:

Let $F, G, H \in S$, i.e. $F(x) = ax + b$, $G(x) = cx + d$, $H(x) = ex + f$, where $a, b, c, d, e, f \in R$ and $a, c, e \neq 0$; then:

G1: $F, G \in S$ implies that $FG(x) = F(G(x)) = a(cx + d) + b = acx + (ad + b) \in S$.

G2: $F(GH(x)) = acex + (acf + ad + b) = FG(H(x))$.

G3: $I(x) = x$, i.e. $r = 1$ and $s = 0$, is the identity of the group.

G4: $M^{-1}(x) = x/r - s/r$.

G5: $FG(x) = acx + (ad + b) \neq acx + (bc + d) = GF(x)$.

Hence the set S is a non-Abelian group.

§3.1.2 Groups of Permutations

Df. 3.1.2.1 A set M of 1-1 transformations R, S, T, \dots on a set E , as defined by Df. 2.2.2.5, forms a *group of transformations* if M satisfies the following conditions:

G1. If $S, T \in M$, then $ST \in M$ is unique.

G3. $TI = IT = T$, $I \in M$.

G4. If $T \in M$, then $T^{-1} \in M$.

Note that **G2** is implicitly contained in the definition of transformation itself (cf. §2.2.2, Prob. 8).

Transformation groups, defined as above, are abstract, but they may become quite concrete, e.g. when exemplified in various arithmetics and geometries (cf. Prob. 1-7). In theory, these groups belong to the group of permutation, defined as below.

Df. 3.1.2.2 A 1-1 transformation on a set E of n distinct elements into E itself is a *permutation* of degree n .

Example:

A set $E = \{1, 2, 3\}$ has a 1-1 transformation T on E into E itself such as $1 \rightarrow 2, 2 \rightarrow 3, 3 \rightarrow 1$, T in this case being the so-called *cyclic permutation*.

The definition of permutation, which is to subsume transformation itself in the framework of groups, may be defined without an explicit use of the term, e.g. as follows:

Df. 3.1.2.2a Given n distinct elements a_1, a_2, \dots, a_n of a set E in this specific (distinct) arrangement, an operation of replacing a_1 by b_1, a_2 by b_2, \dots, a_n by b_n to yield any other (different or same) arrangement b_1, b_2, \dots, b_n of the same n elements, is a *permutation* P , denoted by

$$P = \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ b_1 & b_2 & \dots & b_n \end{pmatrix}$$

to indicate that each element in the first row is to be replaced by the element directly below it in the second row.

It is evidently immaterial here in what order each column of the permutation may be placed, i.e.,

$$P = \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ b_1 & b_2 & \dots & b_n \end{pmatrix} = \begin{pmatrix} a_2 & a_n & \dots & a_1 \\ b_2 & b_n & \dots & b_1 \end{pmatrix} = \begin{pmatrix} a_n & a_1 & \dots & a_2 \\ b_n & b_1 & \dots & b_2 \end{pmatrix} = \dots$$

as long as each permutation remains the same. Since, as elementary algebra proves, there are $n!$ different permutations of n elements, each of the $n!$ permutations can be written in different arrangements, shown as above.

Df. 3.1.2.3 The *product* of two permutations P and Q , denoted by PQ , is obtained by carrying out the operation defined by P and then by Q .

Example:

$$\text{If } P = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} \text{ and } Q = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, \text{ then}$$

$$PQ = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} \begin{pmatrix} 1 & 3 & 2 \\ 2 & 1 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$$

$$\text{and } QP = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \begin{pmatrix} 2 & 3 & 1 \\ 3 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$$

As this example clearly shows, permutations are not always commutative, but they are obviously associative, as can be readily verified by definition itself (cf. Prob. 9). Note, however, that the composite (or product) of transformations, e.g. ST (or RST), defined by Df. 2.2.2.5 and Df. 3.1.2.1, must be written as TS (or TSR) in case transformations are redefined as permutations by Df. 3.1.2.3.

Df. 3.1.2.4 A permutation of degree n which does not affect the product of any number of permutations of the same degree is the *identity permutation*, denoted by I , which obviously has the following general form:

$$I = \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ a_1 & a_2 & \dots & a_n \end{pmatrix}$$

by which $PI = IP = P$ for any permutation P of degree n .

Example:

$$\text{If } I = \begin{pmatrix} 4 & 1 & 3 & 2 \\ 4 & 1 & 3 & 2 \end{pmatrix} \text{ and } P = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix}, \text{ then}$$

$$PI = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix} \begin{pmatrix} 4 & 1 & 3 & 2 \\ 4 & 1 & 3 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix} = P = \begin{pmatrix} 4 & 1 & 3 & 2 \\ 4 & 1 & 3 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix} = IP$$

Df. 3.1.2.5 Iff $PQ = QP = I$, then Q is the *inverse permutation* of P , denoted by P^{-1} .

Inverse transformations in general have the following characteristics:

Th. 3.1.2.6

- (i) Iff $P = \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ b_1 & b_2 & \dots & b_n \end{pmatrix}$, then uniquely $P^{-1} = \begin{pmatrix} b_1 & b_2 & \dots & b_n \\ a_1 & a_2 & \dots & a_n \end{pmatrix}$;
- (ii) $(P^{-1})^{-1} = P$;
- (iii) $(PQ)^{-1} = Q^{-1}P^{-1}$. (Cf. Prob. 10; also Prob. 14).

Th. 3.1.2.7 The set P_n of all permutations on a set S of n elements forms a group of order $n!$ (cf. Prob. 11).

Df. 3.1.2.8 The group P_n of Th. 3.1.2.7 is called the *symmetric permutation* (or *substitution*) *group* of order $n!$ (or degree n), or simply the *symmetric group* of order $n!$, sometimes denoted by S_n instead of P_n .

Df. 3.1.2.9 The group P_n may yield subgroups (cf. Df. 3.1.1.11), which are called *permutation groups* (or *groups of permutations*). P_n being a subgroup of itself, it is also a permutation group.

Example:

If a set S on which permutations operate has only 3 elements, say, a, b, c , then the (symmetric) permutation group P_n is of degree 3 and order $3! = 6$; i.e. the group-forming set P_n has 6 elements, viz.,

$$I = \begin{pmatrix} a & b & c \\ a & b & c \end{pmatrix}, \quad A = \begin{pmatrix} a & b & c \\ a & c & b \end{pmatrix}, \quad B = \begin{pmatrix} a & b & c \\ b & a & c \end{pmatrix},$$

$$C = \begin{pmatrix} a & b & c \\ b & c & a \end{pmatrix}, \quad D = \begin{pmatrix} a & b & c \\ c & a & b \end{pmatrix}, \quad E = \begin{pmatrix} a & b & c \\ c & b & a \end{pmatrix},$$

out of which I, C , and D , for instance, forms a subgroup of P_n , viz. a permutation group of degree 3 and order 3, as can be readily proved (cf. Prob. 6).

Df. 3.1.2.10 A permutation P on a set S of n distinct elements a_1, a_2, \dots, a_n is a *cycle* (or *circular permutation*) of *degree* (or *length*) m if S has a subset S_1 whose m distinct elements b_1, b_2, \dots, b_m , are cyclically interchanged, i.e. $b_1 \rightarrow b_2, b_2 \rightarrow b_3, \dots, b_m \rightarrow b_1$, and $a_i \rightarrow a_i$ for any $a_i \notin S_1$.

Example:

$$P = \begin{pmatrix} a_1 & a_2 & a_3 & \dots & a_{i-1} & a_i & a_{i+1} & \dots & a_n \\ a_1 & a_3 & a_4 & \dots & a_i & a_2 & a_{i+1} & \dots & a_n \end{pmatrix}$$

It is customary, however, that cycles have a contracted form, viz.,

$$P = (a_1)(a_2 a_3 \dots a_i)(a_{i+1}) \dots (a_n)$$

where a cycle consisting of a single element indicates that the element remains unaltered in the new arrangement; such cycles are often entirely omitted, further abbreviating the contracted form, viz.,

$$P = (a_2 a_3 \dots a_i) \quad (= (a_3 \dots a_i a_2) = \dots = (a_i a_2 \dots a_{i-1}))$$

Df. 3.1.2.11 Two or more cycles which have no element in common are called (*mutually disjoint*) (Cf. Prob. 16).

Th. 3.1.2.12 Every permutation can be changed into a product of disjoint cycles. (Cf. Prob. 17.)

Df. 3.1.2.13 In particular, if $m = 1$ in Df. 3.1.1.10, i.e. $a_j \rightarrow a_j$ for any $a_j \in S_1$ (and, as before, $a_i \rightarrow a_i$ for any $a_i \notin S_1$), then $P = (a_1)(a_2) \dots (a_n)$ is the identity permutation I itself, denoted by (1) in this context; if $m = 2$, then a cycle is called a *transposition*.

Th. 3.1.2.14 Every cycle can be expressed as a product of transpositions in infinitely many ways. (Cf. Prob. 18.)

Th. 3.1.2.15 Every permutation can be expressed as a product of transpositions in infinitely many ways. (Cf. Prob. 19.)

Df. 3.1.2.16 A permutation is called *even* if it can be rewritten as a product of an even number of transpositions; otherwise it is called *odd*.

Th. 3.1.2.17 A permutation cannot be both even and odd. (Cf. Prob. 20.)

Df. 3.1.2.18 A complex A_n of even permutations of the symmetric group S_n forms a subgroup of S_n , called the *alternating (sub)group* of S_n .

Example:

$S_2 = \{(1), (12)\}$ has $A_2 = \{(1)\}$, and $S_3 = \{(1), (12), (13), (23), (123), (132)\}$ has $A_3 = \{(1), (123), (132)\}$.

Likewise S_4 has $A_4 = \{(1), (123), (132), (124), (142), (134), (143), (234), (243), (12)(34), (13)(24), (14)(23)\}$.

Th. 3.1.2.19 Of the $n!$ permutations of S_n , A_n consists of $n!/2$ permutations. (Cf. Prob. 21.)

* * * *

In the following pages, to clarify and simplify the process of discovering and examining groups, the so-called *multiplication table* (or *Cayley table*) will be employed; it is a table with a double entry:

	a_1	a_2	\dots	a_j	\dots
a_1	a_{11}	a_{12}	\dots	a_{1j}	\dots
a_2	a_{21}	a_{22}	\dots	a_{2j}	\dots
\vdots	\vdots	\vdots		\vdots	
\vdots	\vdots	\vdots		\vdots	
\vdots	\vdots	\vdots		\vdots	
a_i	a_{i1}	a_{i2}	\dots	a_{ij}	\dots
\vdots	\vdots	\vdots		\vdots	
\vdots	\vdots	\vdots		\vdots	

where $a_{ij} = a_i a_j$.

Example:

A Boolean algebra B (cf. Df. 2.4.2.1), formed by the subsets of $I = \{a, b\}$, may be schematized by the following multiplication table, where $x = \{a\}$, $y = \{b\}$ (and consequently, $x' = y$, $y' = x$, $I' = O$, $O' = I$):

\vee	O	x	y	I
O	O	x	y	I
x	x	x	I	I
y	y	I	y	I
I	I	I	I	I

\wedge	O	x	y	I
O	O	O	O	O
x	O	x	O	x
y	O	O	y	y
I	O	x	y	I

Solved Problems

1. Suppose an owner of a car without a spare tire rotated the tires according to the following patterns:

Rotation 0
F.L. F.R.

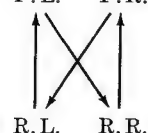
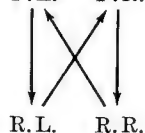
Rotation 1
F.L. \rightleftarrows F.R.

Rotation 2
F.L. F.R.

Rotation 3
F.L. F.R.

R.L. R.R.

R.L. \rightleftarrows R.R.



Show that the rotations form an Abelian group.

PROOF:

The arithmetic of rotating tires is given by the multiplication table at right, where the properties of G1-5 can be quite easily checked.

For example, if xRy represents the binary operation of the first rotation x being followed by the second rotation y , then $1R2 = 3$, $3R2 = 0$, etc., which reveals the closure property (G1). Again, e.g. $1R(2R3) = (1R2)R3 = 1$, etc., affirming the associativity (G2). Since $0R1 = 1R0 = 1$, etc., rotation 0 is obviously the identity (G3). Also $0^{-1} = 0$, $1^{-1} = 1$, $2^{-1} = 3$, $3^{-1} = 2$, providing inverses (G4) for all rotations of tires. Furthermore, e.g. $1R2 = 2R1$, $2R3 = 3R2$, etc., exemplifying G5.

	0	1	2	3
0	0	1	2	3
1	1	0	3	2
2	2	3	1	0
3	3	2	0	1

2. Given the right-hand orthogonal coordinate system (cf. Fig. 3.1.2a), let a, b, c be the clockwise rotations through 180° about X, Y, Z axes respectively, and e be the original position; then e, a, b, c forms an Abelian group.

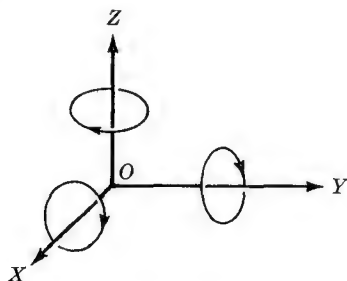


Fig. 3.1.2a

	e	a	b	c
e	e	a	b	c
a	a	e	c	b
b	b	c	e	a
c	c	b	a	e

PROOF:

This set of rotations is characterized by the multiplication table above, which does reveal all of G1-5, when studied as in Prob. 1. The multiplication table, incidentally, reveals a distinguished group, known as the (Klein's) *four group* or *Vierergruppe*, often denoted by V_4 (cf. Prob. 12 below and §3.1.3, Prob. 12).

3. Let an ordered pair of real numbers, (x, y) , 1-1 correspond to a point in the plane. If the point (x, y) is moved horizontally by a units and vertically by b units, it then attains the new position $(x+a, y+b)$, or simply $\{a, b\}$, braces denoting such *translations* in the plane. Now, if the translation $\{a, b\}$ is followed by another $\{c, d\}$, then the total translation is $\{a+c, b+d\}$. Prove that such translations form an Abelian group, called the *group of translations*.

PROOF:

Let T represent the binary operation of translation; then

$$\text{G1: } \{a, b\} T \{c, d\} = \{a+c, b+d\}.$$

$$\begin{aligned} \text{G2: } (\{a, b\} T \{c, d\}) T \{e, f\} &= \{a+c, b+d\} T \{e, f\} = \{a+c+e, b+d+f\} \\ &= \{a+(c+e), b+(d+f)\} = \{a, b\} T \{c+e, d+f\} = \{a, b\} T (\{c, d\} T \{e, f\}). \end{aligned}$$

$$\text{G3: } \{a, b\} T \{0, 0\} = \{a, b\}.$$

$$\text{G4: } \{a, b\} T \{-a, -b\} = \{0, 0\}.$$

$$\text{G5: } \{a, b\} T \{c, d\} = \{a+c, b+d\} = \{c+a, b+d\} = \{c, d\} T \{a, b\}.$$

Hence the translation forms an Abelian group.

4. The rigid motions of an equilateral triangle (cf. Fig. 3.1.2b) entail two sets of *symmetries*: (i) *rotational symmetries* S_0, S_1, S_2 , representing the clockwise rotations through 0° (or 360°), 120° , 240° respectively, and (ii) *reflective symmetries* S_3, S_4, S_5 , representing the reflections in the axes AD, BE, CF respectively. Prove that these symmetries altogether form a group.

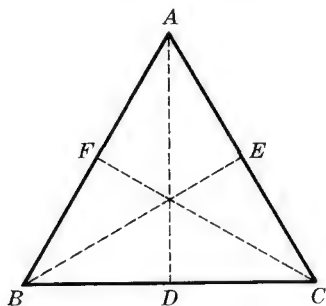
PROOF:

Fig. 3.1.2b

	S_0	S_1	S_2	S_3	S_4	S_5
S_0	S_0	S_1	S_2	S_3	S_4	S_5
S_1	S_1	S_2	S_0	S_5	S_3	S_4
S_2	S_2	S_0	S_1	S_4	S_5	S_3
S_3	S_3	S_4	S_5	S_0	S_1	S_2
S_4	S_4	S_5	S_3	S_2	S_0	S_1
S_5	S_5	S_3	S_4	S_1	S_2	S_0

The multiplication table above does provide all of G1-4 for the set of the rigid motions, which therefore form a group, completing the proof.

Note that, as can be immediately observed, the set of rotational symmetries, itself a group, forms a subgroup of the original group; the latter is not Abelian, but the former is.

5. Repeat the algebra of symmetry, as above in Prob. 4, for the rigid motions of a square, which form a group, called the *dihedral group* (of the square), and denoted sometimes by D_4 . (In general, the dihedral group D_n is the group of all symmetries of a regular polygon of n sides.)

PROOF:

Let 0, 1, 2, 3 be the clockwise rotations through $0^\circ, 90^\circ, 180^\circ, 270^\circ$ respectively and 4, 5, 6, 7 be the reflections in the axes EG, FH, AC, BD (cf. Fig. 3.1.2c) respectively; then, by the following multiplication table, G1-4 can be readily checked.

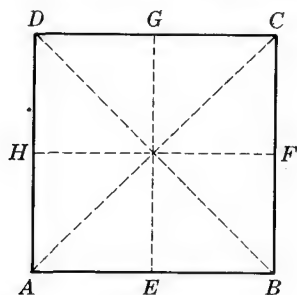
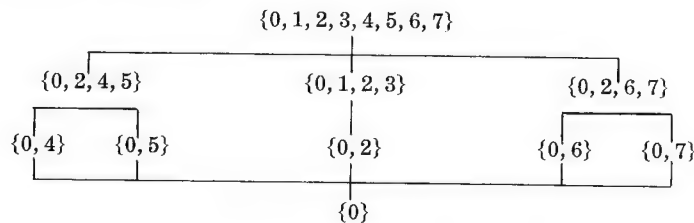


Fig. 3.1.2c

	0	1	2	3	4	5	6	7
0	0	1	2	3	4	5	6	7
1	1	2	3	0	7	6	4	5
2	2	3	0	1	5	4	7	6
3	3	0	1	2	6	7	5	4
4	4	6	5	7	0	2	1	3
5	5	7	4	6	2	0	3	1
6	6	5	7	4	3	1	0	2
7	7	4	6	5	1	3	2	0

Note that here, too, is a subgroup, viz. the set of rotational symmetries, which forms an Abelian group. Note, however, that the subset $\{0, 1, 2, 3\}$ is not the only subgroup of the original group; as a matter of fact, there are eight more, viz. $\{0, 2, 4, 5\}$, $\{0, 2, 6, 7\}$, $\{0, 2\}$, $\{0, 4\}$, $\{0, 5\}$, $\{0, 6\}$, $\{0, 7\}$, $\{0\}$, which may be schematized as follows:



Similar schemata may be obtained without difficulty for the preceding problems.

6. Rewrite the transformations (i.e. rotations and reflections) of Prob. 4-5 in terms of permutations and cycles.

Solution:

- (i) Let 1, 2, 3 represent the vertices of the equilateral triangle; then their transformations are

$$S_0 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix} = (1), \quad S_1 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} = (123), \quad S_2 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} = (132)$$

$$S_3 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} = (23), \quad S_4 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} = (13), \quad S_5 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} = (12)$$

- (ii) Let the vertices of the square A, B, C, D be represented by a, b, c, d ; then their transformations through symmetries are:

$$0 = \begin{pmatrix} a & b & c & d \\ a & b & c & d \end{pmatrix} = (1), \quad 1 = \begin{pmatrix} a & b & c & d \\ b & c & d & a \end{pmatrix} = (abcd), \quad 2 = \begin{pmatrix} a & b & c & d \\ c & d & a & b \end{pmatrix} = (ac)(bd).$$

$$3 = \begin{pmatrix} a & b & c & d \\ d & a & b & c \end{pmatrix} = (adcb), \quad 4 = \begin{pmatrix} a & b & c & d \\ b & a & d & c \end{pmatrix} = (ab)(cd), \quad 5 = \begin{pmatrix} a & b & c & d \\ d & c & b & a \end{pmatrix} = (ad)(bc),$$

$$6 = \begin{pmatrix} a & b & c & d \\ c & b & a & d \end{pmatrix} = (ac), \quad 7 = \begin{pmatrix} a & b & c & d \\ a & d & c & b \end{pmatrix} = (bd)$$

7. Express the group of a regular hexagon in terms of a group of permutations of its vertices.

Solution:

Let $S_0, S_1, S_2, S_3, S_4, S_5$ be the symmetries of rotation through $0^\circ, 60^\circ, 120^\circ, 180^\circ, 240^\circ, 300^\circ$, and $S_6, S_7, S_8, S_9, S_{10}, S_{11}$ represent the symmetries of reflection (cf. Fig. 3.1.2d); then

$$\begin{aligned} S_0 &= (1) \quad (\text{or } (1)(2)(3)(4)(5)(6)) \\ S_1 &= (123456) & S_2 &= (153)(264) \\ S_3 &= (14)(25)(36) & S_4 &= (135)(246) \\ S_5 &= (654321) \\ S_6 &= (16)(25)(34) & S_7 &= (1)(4)(26)(35) \\ S_8 &= (12)(36)(45) & S_9 &= (2)(5)(13)(46) \\ S_{10} &= (14)(23)(56) & S_{11} &= (3)(6)(15)(24) \end{aligned}$$

By completing a multiplication table it can be verified without difficulty that the set of the twelve symmetries actually forms a group (cf. Prob. 12).

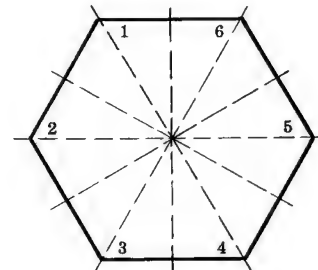


Fig. 3.1.2d

8. Given a regular tetrahedron (cf. Fig. 3.1.2e), find all the symmetries which form a group of rotations.

Solution:

Since the regular tetrahedron has 4 vertices, there exist $4! = 24$ permutations (i.e. all possible rotations, cf. Supplementary Prob. 3.1) but only the following twelve rotations form a group:

$$(1), (ABC), (BDC), (ACB), (AC)(BD), (ADC), (BCD), \\ (AB)(CD), (ABD), (ACD), (ADB), (AD)(BC)$$

as can be readily checked by a multiplication table.

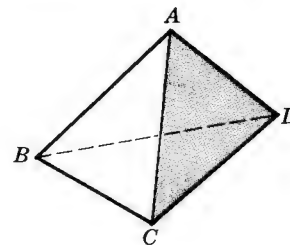


Fig. 3.1.2e

9. Verify the associative law for permutations.

PROOF:

$$\text{Let } P_1 = \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ b_1 & b_2 & \dots & b_n \end{pmatrix}, \quad P_2 = \begin{pmatrix} b_1 & b_2 & \dots & b_n \\ c_1 & c_2 & \dots & c_n \end{pmatrix}, \quad P_3 = \begin{pmatrix} c_1 & c_2 & \dots & c_n \\ d_1 & d_2 & \dots & d_n \end{pmatrix}, \quad \text{where}$$

$b_1 b_2 \dots b_n, c_1 c_2 \dots c_n, d_1 d_2 \dots d_n$ are merely different arrangements of the same n elements $a_1 a_2 \dots a_n$; then

$$P_1 P_2 = \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ c_1 & c_2 & \dots & c_n \end{pmatrix} \quad \text{and} \quad P_2 P_3 = \begin{pmatrix} b_1 & b_2 & \dots & b_n \\ d_1 & d_2 & \dots & d_n \end{pmatrix}$$

$$\text{Hence} \quad (P_1 P_2) P_3 = \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ c_1 & c_2 & \dots & c_n \end{pmatrix} \begin{pmatrix} c_1 & c_2 & \dots & c_n \\ d_1 & d_2 & \dots & d_n \end{pmatrix} = \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ d_1 & d_2 & \dots & d_n \end{pmatrix},$$

and likewise

$$P_1 (P_2 P_3) = \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ d_1 & d_2 & \dots & d_n \end{pmatrix}$$

$$\text{Thus } (P_1 P_2) P_3 = P_1 (P_2 P_3).$$

10. Prove Th. 3.1.2.6.

PROOF:

(i) It is sufficient, since

$$P P^{-1} = \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ b_1 & b_2 & \dots & b_n \end{pmatrix} \begin{pmatrix} b_1 & b_2 & \dots & b_n \\ a_1 & a_2 & \dots & a_n \end{pmatrix} = \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ a_1 & a_2 & \dots & a_n \end{pmatrix} = I$$

$$P^{-1} P = \begin{pmatrix} b_1 & b_2 & \dots & b_n \\ a_1 & a_2 & \dots & a_n \end{pmatrix} \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ b_1 & b_2 & \dots & b_n \end{pmatrix} = \begin{pmatrix} b_1 & b_2 & \dots & b_n \\ b_1 & b_2 & \dots & b_n \end{pmatrix} = I$$

Conversely, it is necessary; for, if $PP^{-1} = PR$ and $P^{-1} \neq R$, then, by Df. 3.1.2.5, $P^{-1}PP^{-1} = P^{-1}PR$, i.e. $IP^{-1} = IR$, i.e. $P^{-1} = R$, which is a contradiction. Hence it must be the case that $R = P^{-1}$. Also, if $P^{-1}P = RP$, then likewise $R = P^{-1}$. Hence P^{-1} is unique.

$$(ii) \quad \text{By (i), } (P^{-1})^{-1} = \begin{pmatrix} b_1 & b_2 & \dots & b_n \\ a_1 & a_2 & \dots & a_n \end{pmatrix}^{-1} = \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ b_1 & b_2 & \dots & b_n \end{pmatrix} = P.$$

(iii) By Problem 9, $(PQ)(Q^{-1}P^{-1}) = P(QQ^{-1})P^{-1} = PIP^{-1} = PP^{-1} = I$. Likewise, $(Q^{-1}P^{-1})(PQ) = I$. Hence, by Df. 3.1.2.5, $(PQ)^{-1} = Q^{-1}P^{-1}$.

11. Prove Th. 3.1.2.7.

PROOF:

G1: $P, Q \in P_n$ implies $PQ \in P_n$, by Df. 3.1.2.3.

G2: $P, Q, R \in P_n$ implies $P(QR) = (PQ)R$, by Prob. 9.

G3: $PI = IP = P$, $I \in P_n$, by Df. 3.1.2.5.

G4: $PP^{-1} = P^{-1}P = I$, by Prob. 10(i).

Hence the set P_n forms a group, satisfying all of G1-4.

12. Prove that the following cycles form a group under permutation:

$$(1), (12)(34), (13)(24), (14)(23)$$

PROOF:

Let $(1) = C_0$, $(12)(34) = C_1$, $(13)(24) = C_2$, $(14)(23) = C_3$. Then their products yield the multiplication table of a group, which also satisfies G5; i.e. it is Abelian.

	C_0	C_1	C_2	C_3
C_0	C_0	C_1	C_2	C_3
C_1	C_1	C_0	C_3	C_2
C_2	C_2	C_3	C_0	C_1
C_3	C_3	C_2	C_1	C_0

13. Find a group of permutations on a set $\{1, 2, 3, 4\}$ for which a mapping

$$f(x_1, x_2, x_3, x_4) = x_1x_2 + x_3x_4$$

remains invariant.

Solution:

Let π be the operator of the prescribed permutations; i.e. $\pi = \begin{pmatrix} 1 & 2 & 3 & 4 \\ a & b & c & d \end{pmatrix}$ implies $\pi f = x_ax_b + x_cx_d$.

Now, if the mapping is to remain unvaried under the permutation, i.e. $\pi f = f$, then it must be the case that, e.g. (i) $x_ax_b = x_1x_2$, $x_cx_d = x_3x_4$, (ii) $x_ax_b = x_3x_4$, $x_cx_d = x_1x_2$, etc. Such permutations are

$$(1), (12), (34), (12)(34), (13)(24), (14)(23), (1324), (1423)$$

which form a group, as can be verified without difficulty (cf. Prob. 12 above).

14. Prove $(a_1 a_2 \dots a_n)^{-1} = (a_n \dots a_2 a_1)$.

PROOF:

Since

$$\begin{aligned} (a_1 a_2 \dots a_n)(a_n \dots a_2 a_1) &= \begin{pmatrix} a_1 & a_2 & \dots & a_{n-1} & a_n \\ a_2 & a_3 & \dots & a_n & a_1 \end{pmatrix} \begin{pmatrix} a_n & \dots & a_2 & a_1 \\ a_{n-1} & \dots & a_1 & a_n \end{pmatrix} \\ &= \begin{pmatrix} a_1 & a_2 & \dots & a_{n-1} & a_n \\ a_2 & a_3 & \dots & a_n & a_1 \end{pmatrix} \begin{pmatrix} a_2 & a_3 & \dots & a_n & a_1 \\ a_1 & a_2 & \dots & a_{n-1} & a_n \end{pmatrix} = \begin{pmatrix} a_1 & a_2 & \dots & a_{n-1} & a_n \\ a_1 & a_2 & \dots & a_{n-1} & a_n \end{pmatrix} \\ &= (1) \end{aligned}$$

and likewise $(a_n \dots a_2 a_1)(a_1 a_2 \dots a_n) = (1)$, then by Df. 3.1.2.5, $(a_1 a_2 \dots a_n)^{-1} = (a_n \dots a_2 a_1)$.

15. Prove: (i) $(a_1 a_2 \dots a_n)(a_1 a_2)(a_n \dots a_2 a_1) = (a_1 a_n)$
 (ii) $(a_1 a_2 \dots a_n)(a_1 a_{n+1}) = (a_1 a_2 \dots a_n a_{n+1})$

PROOF:

$$\begin{aligned} \text{(i)} \quad (a_1 a_2 \dots a_n)(a_1 a_2)(a_n \dots a_2 a_1) &= \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ a_2 & a_3 & \dots & a_1 \end{pmatrix} \begin{pmatrix} a_1 & a_2 & a_3 & \dots & a_n \\ a_2 & a_1 & a_3 & \dots & a_n \end{pmatrix} \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ a_n & a_1 & \dots & a_{n-1} \end{pmatrix} \\ &= \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ a_n & a_2 & \dots & a_1 \end{pmatrix} = (a_1 a_n) \end{aligned}$$

$$\begin{aligned} \text{(ii)} \quad (a_1 a_2 \dots a_n)(a_1 a_{n+1}) &= \begin{pmatrix} a_1 & a_2 & \dots & a_n & a_{n+1} \\ a_2 & a_3 & \dots & a_1 & a_{n+1} \end{pmatrix} \begin{pmatrix} a_1 & a_2 & \dots & a_n & a_{n+1} \\ a_{n+1} & a_2 & \dots & a_n & a_1 \end{pmatrix} \\ &= \begin{pmatrix} a_1 & a_2 & \dots & a_n & a_{n+1} \\ a_2 & a_3 & \dots & a_{n+1} & a_1 \end{pmatrix} = (a_1 a_2 \dots a_n a_{n+1}) \end{aligned}$$

16. Verify (i) $(123)(45) = (45)(123)$ and (ii) $(123)(23) \neq (23)(123)$; then generalize, i.e. prove that $PQ = QP$ if P and Q are disjoint cycles.

PROOF:

$$\text{(i)} \quad (123)(45) = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 1 & 4 & 5 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 3 & 5 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 1 & 5 & 4 \end{pmatrix} = \begin{pmatrix} 4 & 5 & 1 & 2 & 3 \\ 5 & 4 & 2 & 3 & 1 \end{pmatrix} = (45)(123).$$

$$\text{(ii)} \quad (123)(23) = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} = (13),$$

$$(23)(123) = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} = (12) \neq (13) = (123)(23).$$

In general, therefore, $PQ = QP$ if P and Q are disjoint, since the elements permuted by P are left unchanged by Q and also the elements permuted by Q remain the same under P .

17. Prove Th. 3.1.2.12.

PROOF:

$$\text{Let } P = \begin{pmatrix} a_1 & a_2 & \dots & a_i & \dots & a_k & \dots & a_n \\ b_1 & b_2 & \dots & b_i & \dots & b_k & \dots & b_n \end{pmatrix}.$$

- (i) If the elements in the first row of P are all completely different from the elements in the second row of P , it is then obviously the case that $P = (a_1 b_1) \dots (a_i b_i) \dots (a_n b_n)$, proving the theorem.
- (ii) If the first row has some elements in common with the second row, and if $b_1 \neq a_1$, then b_1 may be found in the first row of P . Suppose $b_1 = a_i$, then $b_1 \neq b_i$, since b_1 cannot occur twice in the second row (i.e. all elements in each row are distinct, by Df. 3.1.2.2). Now, if $b_i = a_1$, then the cycle $(a_1 b_i)$ is closed.

If $b_i \neq a_1$, then b_i may be found in the first row of P , and suppose $b_i = a_k$; then b_k is different from b_1 and b_i . Now, if $b_k = a_1$, then the cycle $(a_1 b_i b_k)$ is closed.

If $b_k \neq a_1$, then the process can be continued until, after at most n steps, the cycle closes.

If the first cycle thus obtained does not involve all the elements of P , then take b_2 as any other element in the first row of P and, repeating the process prescribed as above, another cycle can be obtained. If the process is reiterated until the elements of P are exhausted, P is then expressed as a product of disjoint cycles, establishing the theorem in general.

Note. In particular, the identity permutation on n elements is expressed as (1) or n cycles each of which has length 1, viz. $(a_1)(a_2) \dots (a_n)$. (Cf. Df. 3.1.2.13.)

How readily the process works and simplifies the matter has already been tested by Prob. 6, etc.

21. Prove Th. 3.1.2.19.**PROOF:**

Let P_n of order $n!$ consist of r even permutations p_1, p_2, \dots, p_r and s odd permutations q_1, q_2, \dots, q_s (and nothing else, i.e. $r + s = n!$, cf. Th. 3.1.2.17). Now multiply p_1, p_2, \dots, p_r by a transposition t , i.e. tp_1, tp_2, \dots, tp_r , which are r odd permutations and all distinct, since if, e.g., $tp_i = tp_j$, then $p_i = (t^{-1}t)p_i = t^{-1}(tp_i) = t^{-1}(tp_j) = (t^{-1}t)p_j = p_j$, contrary to the assumption. Hence the odd permutations of the form tp_k , $k = 1, 2, \dots, r$, must be all distinct and consequently $s \leq r$.

Likewise, multiplying q_1, q_2, \dots, q_s by t , which now produces s distinct even permutations, $r \leq s$.

Hence $r = s$, and since $r + s = n!$, it follows that $r = s = n!/2$.

§3.1.3 Homomorphism and Isomorphism

Th. 3.1.3.1 Given an $(X, *; Y, \circ)$ -homomorphism f of a set X onto or into a set Y , i.e. $f(a * b) = f(a) \circ f(b)$ for all $a, b \in X$ (cf. Df. 2.2.2.9), it follows that

- (i) \circ is associative if $*$ is associative,
- (ii) \circ is commutative if $*$ is commutative, and
- (iii) $e \in X$, which is an identity under $*$, corresponds to an identity under \circ , $f(e) \in Y$. (Cf. Prob. 1).

These properties belong to the homomorphism in general of one set onto the other. The corresponding *sets* in this context may be replaced by a pair of corresponding *groups* through the following definition and theorem.

Df. 3.1.3.2 A *homomorphism* of a group G onto or into a group G' is a transformation H of G onto or into all of G' such that, for all $x, y \in G$,

$$H(xy) = H(x)H(y) = x'y'$$

for all $x', y' \in G'$. $G' = H(G)$ itself is called here a *homomorph* (i.e. homomorphic image) of G .

Note that the operator which operates on x and y may be different from the operator which operates on x' and y' (cf. Prob. 4). Note, also, that the correspondence in a homomorphism may be *many-one* (cf. Prob. 5-8).

Th. 3.1.3.3 If, in addition to (i)-(iii) in Th. 3.1.3.1, there exists

- (iv) $f(a^{-1}) \in Y$ such that $f(a) \circ f(a^{-1}) = f(e)$, corresponding to $a^{-1} \in X$ such that $a * a^{-1} = e$,

then it is an $(X, *; Y, \circ)$ -homomorphism f of a group X onto or into a group Y . (Cf. Prob. 2.)

And conversely, as in the following theorem:

Th. 3.1.3.4 If X is a group under $*$, and if f is an $(X, *; Y, \circ)$ -homomorphism of X onto or into Y , then Y is a group under \circ . (Cf. Prob. 3.)

Th. 3.1.3.5 The relation of homomorphism is transitive (cf. Df. 2.1.1.12); i.e. if G is homomorphic to G' which in turn is homomorphic to G'' , then G is homomorphic to G'' . (Cf. Prob. 4.)

Df. 3.1.3.6 If a $(G, *; G', \circ)$ -homomorphism of G onto or into all of G' is one-one, it is then a $(G, *; G', \circ)$ -isomorphism of G into (or more precisely, onto and into) G' .

Stated otherwise: If there exists a 1-1 transformation f of a group G into a group G' such that

- (i) there exists a unique correspondent $f(x) \in G'$ for every $x \in G$, and
- (ii) there exists a unique correspondent $x \in G$ for which $f(x) \in G'$,

then the 1-1 correspondence is an isomorphism between G and G' .

Notationally, the isomorphism may be easily distinguished from the homomorphism by the use of two-way, against one-way, arrows, since there does exist a two-way traffic in isomorphisms while such a traffic, by definition, is not always assured for homomorphisms. As such, Df. 3.1.3.2 and Df. 3.1.3.6 may be put together as follows:

A homomorphism of G onto or into G' is a mapping $G \rightarrow G'$ such that $x \rightarrow x'$ and $y \rightarrow y'$ imply $(xy)' \rightarrow x'y'$ for all $x, y \in G$, and an isomorphism of G onto and into G' is a one-one mapping $G \leftrightarrow G'$ such that $x \leftrightarrow x'$ and $y \leftrightarrow y'$ imply $xy \leftrightarrow x'y'$.

Th. 3.1.3.7 The relation of isomorphism is reflexive, symmetric, and transitive (cf. Df. 2.1.1.12 and Th. 3.1.3.5). (Cf. Prob. 9 below.)

This relation obviously does not hold for a homomorphism of G onto or into G' where the elements of G' may satisfy additional properties which cannot be found in G . The following theorem of isomorphisms, however, is similar to Th. 3.1.3.4 of homomorphisms.

Th. 3.1.3.8 If G' is a set closed with respect to \circ , and if there exists a $(G, *; G', \circ)$ -isomorphism between G , a group under $*$, and G' , then G' is a group under \circ . (Cf. Prob. 10.)

This theorem may be employed to prove an unidentified structure to be a group. Also, as in Th. 3.1.3.4, it has a converse form:

Th. 3.1.3.9 If there exists an isomorphism between G and G' , then the identities of G and G' correspond and the inverses of corresponding elements in G and G' correspond. (Cf. Prob. 10.)

The importance of the concept of isomorphism is embodied, for instance, in the following theorem (by Cayley), which also may be interpreted as assuring the completeness of G1-4 with respect to transformations, i.e. permutations.

Th. 3.1.3.10 (by Cayley). There exists an isomorphism between any abstract group G of order n and a suitable group of permutations P_n of degree n (i.e. order $n!$) (Cf. Prob. 18.)

In general, then, the properties of any finite abstract group G will be immediately found by examining the properties of a permutation group P which is isomorphic to G ; this is an explicit advantage, since many properties of permutation groups are readily available through detailed studies in the past.

Solved Problems

1. Prove Th. 3.1.3.1.

PROOF:

If $a, b, c \in X$, then, by Df. 2.2.2.9, $a \rightarrow f(a)$, $b \rightarrow f(b)$, $c \rightarrow f(c)$, where $f(a), f(b), f(c) \in Y$. Using the same definition and the given hypotheses,

- (i) $f(a * (b * c)) = f(a) \circ f(b * c) = f(a) \circ (f(b) \circ f(c))$, and $f((a * b) * c) = f(a * b) \circ f(c) = (f(a) \circ f(b)) \circ f(c)$. But, since $a * (b * c) = (a * b) * c$, it directly follows that $f(a * (b * c)) = f((a * b) * c)$, and consequently that $f(a) \circ (f(b) \circ f(c)) = (f(a) \circ f(b)) \circ f(c)$.
- (ii) can be proved likewise.
- (iii) Since $a * e = e * a = a$ for all $a \in X$, $f(a * e) = f(e * a) = f(a)$, i.e. $f(a) \circ f(e) = f(e) \circ f(a) = f(a)$. But $a \rightarrow f(a)$ and $a \in X$ implies $f(a) \in Y$; hence $f(e) \in Y$, which then is the identity of Y under \circ .

2. Prove Th. 3.1.3.3.

PROOF:

G1: By Df. 2.2.2.9, $a * b \in X$ does imply $f(a) \circ f(b) \in Y$.

G2-3: Proved by Prob. 1(i), (iii).

G4: Given here as a hypothesis.

Hence, satisfying G1-4, both X and Y are groups and the homomorphism is of one group onto or into the other.

3. Prove Th. 3.1.3.4.

PROOF:

Since X is already a group under $*$, there exists $a^{-1} \in X$ for all $a \in X$ such that $a * a^{-1} = a^{-1} * a = e$, where $e \in X$. And, by the prescribed homomorphism, $f(a) \circ f(a^{-1}) = f(a^{-1}) \circ f(a) = f(e)$, where $f(e)$ is the identity of Y , for all $f(a) \in Y$. Hence the inverse in Y of $f(a)$ does exist, which is $f(a^{-1})$, providing G4 for the given set Y .

Since Y has already satisfied G1-3 through Df. 2.2.2.9 and Th. 3.1.3.1, it is now proved to be a group under \circ .

4. Prove Th. 3.1.3.5.

PROOF:

If G, G', G'' are all groups, it then follows, directly from Th. 3.1.3.4, that $G \rightarrow G'$ and $G' \rightarrow G''$ imply $G \rightarrow G''$.

In general, let S be the homomorphism of G onto or into G' , and T the homomorphism of G' onto or into G'' . Then, by S , $a \in G \rightarrow a' \in G'$ and $b \in G \rightarrow b' \in G'$ imply $ab \in G \rightarrow a'b' \in G'$, and by T , $a' \in G' \rightarrow a'' \in G''$ and $b' \in G' \rightarrow b'' \in G''$ imply $a'b' \in G' \rightarrow a''b'' \in G''$. Hence, by S and T , $a \in G \rightarrow a'' \in G''$ and $b \in G \rightarrow b'' \in G''$ imply $ab \in G \rightarrow a''b'' \in G''$, i.e. there exists a homomorphism of G onto or into G'' .

5. Let I_+ be the additive group of all integers; then a mapping f , defined by $f(n) = 2n$, is a homomorphism of I_+ into I_+ under addition.

PROOF:

By hypothesis, $n_1 \rightarrow f(n_1) = 2n_1$ and $n_2 \rightarrow f(n_2) = 2n_2$, which together imply, by hypothesis,

$$n_1 + n_2 \rightarrow f(n_1 + n_2) = 2(n_1 + n_2) = 2n_1 + 2n_2 = f(n_1) + f(n_2)$$

proving that f is a homomorphism of I_+ into I_+ under addition.

6. Verify that a set $S = \{1, -1\}$ forms a multiplicative group; then prove that there exists a homomorphism of P_n , the symmetric group of degree n , onto S .

PROOF:

S does form a group under multiplication, as can be immediately verified by the multiplication table shown below.

Now, let π be the operator defined by Th. 3.1.2.17 (cf. §3.1.2, Prob. 20); then $\pi P_n = 1$ if P_n is an even permutation, and $\pi P_n = -1$ if P_n is an odd permutation. Also $\pi P = 1 \rightarrow 1$ and $\pi P = -1 \rightarrow -1$; thus $(\pi P = 1)(\pi P = -1) = (\pi P = -1) \rightarrow (1)(-1) = -1$, which is obviously a many-one correspondence. Hence π is a homomorphism of P_n onto S , both under multiplication.

\times	1	-1
1	1	-1
-1	-1	1

7. Verify that the set E of the four roots of $x^4 - 1 = 0$, i.e. $E = \{1, -1, i, -i\}$, forms a multiplicative group, then prove that a transformation T , $T(n) = i^n$, is a homomorphism of I_+ (cf. Prob. 5 above) under addition onto E under multiplication.

PROOF:

E is a multiplicative group, as can be proved by the multiplication table at right, and I_+ is an additive group. Since, by hypothesis, $p \rightarrow T(p) = i^p$ and $q \rightarrow T(q) = i^q$ for all $p, q \in I_+$, it follows that

$$p + q \rightarrow T(p + q) = i^{p+q} = i^p i^q = T(p)T(q)$$

which definitely is a many-one correspondence, proving that T is a homomorphism of I_+ under addition onto E under multiplication.

\times	1	-1	i	$-i$
1	1	-1	i	$-i$
-1	-1	1	$-i$	i
i	i	$-i$	-1	1
$-i$	$-i$	i	1	-1

8. Let $S = \{S_0, S_1, S_2, \dots, S_{11}\}$ be the group of a regular hexagon obtained by six rotations and six reflections (cf. §3.1.2, Prob. 7 and Fig. 3.1.2d) and $R = \{R_0, R_1, R_2\}$ be a set of axes, R_0, R_1, R_2 representing AD, BE, CF respectively, in which vertices are reflected. Find a homomorphism with respect to R and S .

Solution:

Let $(R_0 R_1)$, for instance, represent the permutation of vertices by reflecting in R_0 , then in R_1 ; S is then expressed in terms of the permutation of R as follows:

$$\begin{array}{ll} S_0 \rightarrow (1) & S_6 \rightarrow (R_0 R_1) \\ S_1 \rightarrow (R_0 R_1 R_2) & S_7 \rightarrow (R_1 R_2) \\ S_2 \rightarrow (R_0 R_2 R_1) & S_8 \rightarrow (R_0 R_1 R_2)(R_1 R_2) = (R_0 R_2) \\ S_3 \rightarrow (1) & S_9 \rightarrow (R_0 R_2) \\ S_4 \rightarrow (R_0 R_1 R_2) & S_{10} \rightarrow (R_1 R_2) \\ S_5 \rightarrow (R_0 R_2 R_1) & S_{11} \rightarrow (R_0 R_2 R_1)(R_1 R_2) = (R_0 R_1) \end{array}$$

i.e. there are six two-one correspondences, viz.,

$$\begin{array}{ll} S_0, S_3 \rightarrow (1) & S_1, S_4 \rightarrow (R_0 R_1 R_2) \\ S_2, S_6 \rightarrow (R_0 R_2 R_1) & S_6, S_{11} \rightarrow (R_0 R_1) \\ S_7, S_{10} \rightarrow (R_1 R_2) & S_8, S_9 \rightarrow (R_0 R_2) \end{array}$$

Moreover, e.g. $S_1 S_4 = S_5 \rightarrow (R_0 R_1 R_2)(R_0 R_1 R_2) = (R_0 R_2 R_1)$, as prescribed by hypothesis; this proves a homomorphism of S onto R , both under permutation.

9. Prove Th. 3.1.3.7.

PROOF:

- (i) It follows, directly from Df. 3.1.3.6, that $G \leftrightarrow G$, since $a \in G \leftrightarrow a \in G$ and $b \in G \leftrightarrow b \in G$ imply $ab \in G \leftrightarrow ab \in G$, i.e. $G \leftrightarrow G$. Likewise $G' \leftrightarrow G'$ and $G'' \leftrightarrow G''$.
- (ii) Since $a \in G \leftrightarrow a' \in G'$, $b \in G \leftrightarrow b' \in G'$, and $ab \in G \leftrightarrow a'b' \in G'$ imply $a' \in G' \leftrightarrow a \in G$, $b' \in G' \leftrightarrow b \in G$, and $a'b' \in G' \leftrightarrow ab \in G$, it immediately follows that $G \leftrightarrow G'$ implies $G' \leftrightarrow G$. Likewise $G' \leftrightarrow G''$ implies $G'' \leftrightarrow G'$, and $G'' \leftrightarrow G$ implies $G \leftrightarrow G''$.
- (iii) $a \in G \leftrightarrow a' \in G'$ and $a' \in G \leftrightarrow a'' \in G''$ imply, by Df. 2.1.1.12, that $a \in G \leftrightarrow a'' \in G''$. Likewise $b \in G \leftrightarrow b' \in G'$ and $b' \in G' \leftrightarrow b'' \in G''$ imply $b \in G \leftrightarrow b'' \in G''$. Hence, by Df. 3.1.3.5, $a \in G \leftrightarrow a'' \in G''$ and $b \in G \leftrightarrow b'' \in G''$ imply $ab \in G \leftrightarrow a''b'' \in G''$, i.e. $G \leftrightarrow G''$.

Isomorphism is thus reflexive, symmetric, and transitive.

10. Prove Th. 3.1.3.9.

PROOF:

- (i) $a \in G \leftrightarrow a' \in G'$ and $e \in G \leftrightarrow e' \in G'$ imply, by Th. 3.1.3.7, that $ae \in G \leftrightarrow a'e' \in G'$ and $ea \in G \leftrightarrow e'a' \in G'$. But $ae = a$ and $ea = a$ in G . Hence $a'e' = a'$ and $e'a' = a'$ in G' , proving that e' is the identity element of G' .
- (ii) $a \in G \leftrightarrow a' \in G'$ and $a^{-1} \in G \leftrightarrow a'^{-1} \in G'$ imply that $aa^{-1} = a^{-1}a \in G \leftrightarrow a'a'^{-1} = a'^{-1}a' \in G'$, which in turn implies $aa^{-1} = a^{-1}a = e \in G \leftrightarrow a'a'^{-1} = a'^{-1}a' = e' \in G'$. Hence a'^{-1} is the inverse of a' in G' .

11. Find an isomorphism between the group E of four roots (cf. Prob. 7 above) and the group T of tire rotations (cf. §3.1.2, Prob. 1).**Solution:**

Observe the four 1-1 correspondences between E and T : $1 \leftrightarrow 0$, $-1 \leftrightarrow 1$, $i \leftrightarrow 2$, $-i \leftrightarrow 3$. Also, in general, if $x, y \in E$ and $x', y' \in T$, then $x \leftrightarrow x'$ and $y \leftrightarrow y'$ imply $xy \leftrightarrow x'y'$. E and T are thus isomorphic.

12. Find a group R of symmetries of the rectangle (cf. Fig. 3.1.3a below); then establish an isomorphism between R and the four group V_4 .**Solution:**

Since a 180° rotation yields a transformation B of the vertices 1, 2, 3, 4: $B = (13)(24)$, and since there are two reflective symmetries: $C = (12)(34)$ and $D = (14)(23)$, R does have four elements: A, B, C, D , where A is the identity transformation. R also forms an Abelian group, as can be easily verified by a multiplication table (cf. §3.1.2, Prob. 12) where, in comparison with the multiplication table of V_4 , the following four 1-1 correspondences can be observed: $A \leftrightarrow e$, $B \leftrightarrow a$, $C \leftrightarrow b$, $D \leftrightarrow c$. Hence R and V_4 are isomorphic.

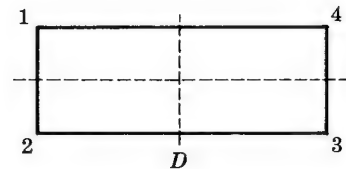


Fig. 3.1.3a

13. Prove that the additive group A of all real numbers is isomorphic to the multiplicative group M of all positive real numbers, excluding 0.**PROOF:**

Since $10^a = 10^b$ iff $a = b$, where $a, b \in A$, and since 10^a , $a \in A$, has the unique inverse $\log_{10} 10^a$, there follow $a \in A \leftrightarrow 10^a \in M$ and $b \in A \leftrightarrow 10^b \in M$ which together imply $(a + b) \in A \leftrightarrow 10^{a+b} = (10^a)(10^b) \in M$. Hence there exists an isomorphism between A under addition and M under multiplication.

14. There exists an isomorphism between the transformation group S of an equilateral triangle with respect to symmetries (cf. §3.1.2, Prob. 4) and the permutation group P of degree 3.**PROOF:**

P is of order $3! = 6$ and has thus 6 elements: $(1), (12), (13), (23), (123), (132)$, which are put into 1-1 correspondence with the six members of S as follows (cf. §3.1.2, Prob. 6):

$$(1) \leftrightarrow S_0, \quad (12) \leftrightarrow S_5, \quad (13) \leftrightarrow S_4, \quad (23) \leftrightarrow S_3, \quad (123) \leftrightarrow S_1, \quad (132) \leftrightarrow S_2$$

Hence S is isomorphic to P .

15. Generalize Prob. 14, i.e. prove that any rotational group R of a regular polygon (or polyhedron) is isomorphic to a suitable permutation group P .**PROOF:**

Let the vertices of the regular polygon in general be a_1, a_2, \dots, a_n , which are first mapped onto a_2, a_3, \dots, a_1 by a rotation through $360^\circ/n$. Represent the first transformation R_1 by a permutation P_1 , viz., $a_1 = P_1(a_n), a_2 = P_1(a_1), \dots, a_n = P_1(a_{n-1})$, i.e.,

$$R_1 \leftrightarrow P_1 = \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ a_2 & a_3 & \dots & a_1 \end{pmatrix}, \quad \text{and likewise}$$

$$R_2 \leftrightarrow P_2 = \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ a_3 & a_4 & \dots & a_2 \end{pmatrix}, \quad \text{and in general}$$

$$R_i \leftrightarrow P_i = \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ a_{i+1} & a_{i+2} & \dots & a_i \end{pmatrix}$$

And, as usual, let the rotation through 0° be

$$R_0 \leftrightarrow P_0 = \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ a_1 & a_2 & \dots & a_n \end{pmatrix}$$

Then, since it obviously holds that $R_1 R_2 \leftrightarrow P_1 P_2$ or in general $R_i R_{i+1} \leftrightarrow P_i P_{i+1}$, and since the set $R = \{R_0, R_1, \dots, R_{n-1}\}$ forms a group, it follows from Th. 3.1.3.8 that the set $P = \{P_0, P_1, \dots, P_{n-1}\}$, which is isomorphic to R , also forms a group.

16. If a group G has a multiplicative rule: $a^2 = b^3 = e$ and $ab = b^2a$ for every $a, b \in G$, then G is isomorphic to S_3 .

PROOF:

Since $e, a, b, b^2, b^2a = ab$ are evidently distinct and $ab^2 = b^2ab = b^2b^2a = b^3ba = ba$, G is of degree 3 (e, a, b) and of order 6, which corresponds to the degree and order of S_3 . Moreover, there exist three basic 1-1 correspondences:

$$e \leftrightarrow (1), \quad a \leftrightarrow (12), \quad b \leftrightarrow (123)$$

which yield three other distinct 1-1 correspondences with respect to S_3 :

$$ab \leftrightarrow (13), \quad ba \leftrightarrow (23), \quad b^2 \leftrightarrow (132)$$

completing the proof.

17. Verify that the four group V_4 , or any (abstract) group which is isomorphic to V_4 , is isomorphic to a suitable permutation group P .

PROOF:

It follows from the multiplication table of V_4 (cf. §3.1.2, Prob. 2) that

$$e \leftrightarrow \begin{pmatrix} e & a & b & c \\ e & a & b & c \end{pmatrix} = (1)$$

$$a \leftrightarrow \begin{pmatrix} e & a & b & c \\ ea & aa & ba & ca \end{pmatrix} = \begin{pmatrix} e & a & b & c \\ a & e & c & b \end{pmatrix} = (ea)(bc)$$

$$b \leftrightarrow \begin{pmatrix} e & a & b & c \\ eb & ab & bb & cb \end{pmatrix} = \begin{pmatrix} e & a & b & c \\ b & c & e & a \end{pmatrix} = (eb)(ac)$$

$$c \leftrightarrow \begin{pmatrix} e & a & b & c \\ ec & ac & bc & cc \end{pmatrix} = \begin{pmatrix} e & a & b & c \\ c & b & a & e \end{pmatrix} = (ec)(ab), \quad \text{and that, e.g.,}$$

$$\begin{aligned} ab \leftrightarrow \begin{pmatrix} e & a & b & c \\ eab & aab & bab & cab \end{pmatrix} &= \begin{pmatrix} e & a & b & c \\ ea & aa & ba & ca \end{pmatrix} \begin{pmatrix} ea & aa & ba & ca \\ eab & aab & bab & cab \end{pmatrix} \\ &= \begin{pmatrix} e & a & b & c \\ ea & aa & ba & ca \end{pmatrix} \begin{pmatrix} e & a & b & c \\ eb & ab & bb & cb \end{pmatrix} = ((ea)(ab))((eb)(ac)) \end{aligned}$$

Hence the four group $V_4 = \{e, a, b, c\}$ is isomorphic to the permutation group

$$P = \{(1), (ea)(ab), (eb)(ac), (ec)(ab)\}$$

18. Generalize Prob. 17, i.e. prove Th. 3.1.3.10.

PROOF:

Let the general abstract group of order n be the set $G = \{G_1, G_2, \dots, G_n\}$; then let each element of G 1-1 correspond to an element of the set P of permutations of degree n , $P = \{P_1, P_2, \dots, P_n\}$, such that

$$G_r \leftrightarrow P_r \equiv \begin{pmatrix} a_i \\ a_i p_r \end{pmatrix} \equiv \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ a_1 p_r & a_2 p_r & \dots & a_n p_r \end{pmatrix}$$

and

$$G_s \leftrightarrow P_s \equiv \begin{pmatrix} a_i \\ a_i p_s \end{pmatrix} \equiv \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ a_1 p_s & a_2 p_s & \dots & a_n p_s \end{pmatrix}$$

where $a_i p_r$ or $a_i p_s$ is obviously the transformation of a_i through the permutation P_r or P_s , i.e. $P_r(a_i) = a_i p_r$ or $P_s(a_i) = a_i p_s$; hence

$$\begin{aligned} G_r G_s \leftrightarrow P_r P_s &\equiv \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ a_1 p_r & a_2 p_r & \dots & a_n p_r \end{pmatrix} \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ a_1 p_s & a_2 p_s & \dots & a_n p_s \end{pmatrix} \\ &= \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ a_1 p_r p_s & a_2 p_r p_s & \dots & a_n p_r p_s \end{pmatrix} = \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ a_1 p_t & a_2 p_t & \dots & a_n p_t \end{pmatrix} \end{aligned}$$

where $a_i p_t \equiv a_i p_r p_s \equiv P_s(P_r(a_i))$ which, by the definition of permutations (and also of transformations (cf. §2.2.2, Prob. 7)), belongs to the original set of permutations P . That is, in general, $G_r \in G \leftrightarrow P_r \in P$ and $G_s \in G \leftrightarrow P_s \in P$ imply

$$G_r G_s \in G \leftrightarrow P_r P_s \equiv \begin{pmatrix} a_i \\ a_i p_r \end{pmatrix} \begin{pmatrix} a_i \\ a_i p_s \end{pmatrix} = \begin{pmatrix} a_i \\ a_i p_r p_s \end{pmatrix} = \begin{pmatrix} a_i \\ a_i p_t \end{pmatrix} \equiv P_t \in P$$

Then, since G is a group and P is now isomorphic to G , P is also a group by Th. 3.1.3.8, completing the proof.

Chapter 3.2

Subgroups

§3.2.1 Cyclic Subgroups

Th. 3.2.1.1 A complex (i.e. non-empty subset) S of a group G is a *subgroup* (cf. Df. 3.1.1.11) iff S satisfies **G1**, i.e. iff $ab \in S$ for every $a, b \in S$. (Cf. Prob. 1 below.)

Example:

D_4 (cf. Prob. 5 of §3.1.2) has ten subgroups, each of which does satisfy Th. 3.2.1.1.

Th. 3.2.1.2 Any complex S of a group G is a subgroup iff $a \in S$ and $b \in S$ imply $a^{-1}b \in S$. (Cf. Prob. 2.)

Example:

The ten subgroups of D_4 .

Note. Th. 3.2.1.2, as well as Th. 3.2.1.1, is actually a simplified version of Th. 3.1.1.13.

Th. 3.2.1.3 If A and B are two subgroups of a group G , then their meet, $A \cap B$, which is called the *common subgroup* of A and B , is also a subgroup of G . (Cf. Prob. 4.)

Example:

$\{0, 1, 2, 3\}$ and $\{0, 2, 4, 5\}$ being the subgroups of D_4 , their meet $\{0, 2\}$ does form a subgroup of D_4 ; or, likewise, $\{0\}$ is a subgroup of D_4 on the strength of its being the meet of $\{0, 2\}$ and $\{0, 4\}$, for instance.

Th. 3.2.1.4 If G is a group and $a \in G$ where $a \neq e$, then a set C of elements a^n , where $n \in I$ (the set of all integers), forms a group. (Cf. Prob. 5.)

Example:

Cf. Prob. 6, 11, etc., below.

Df. 3.2.1.5 The group C of Th. 3.2.1.4 is called the *cyclic subgroup* generated by a , and a itself is called a *generator* of C . (Cf. Prob. 7, 9, 11, 12, etc.)

Df. 3.2.1.6 The *order* of an element $a \in G$ is the order of the cyclic subgroup C , $C \subseteq G$, and is the smallest positive number n for which $a^n = e$; if $a^n \neq e$ for any $n \in I$, then the order of a (or C itself) is said to be infinite. (Cf. Prob. 6, 7, 9, etc.)

Note that the existence or non-existence of n is guaranteed by the Well-ordering Principle (cf. Df. 2.2.1.10), and that e itself is the only element of G which is of order 1.

Th. 3.2.1.7 Any cyclic group is Abelian. (Cf. Prob. 16.)

The problem of determining all subgroups of a specific group is generally complicated, but not for cyclic groups, which are taken care of by the following theorem.

Th. 3.2.1.8 All subgroups of a cyclic group are cyclic. (Cf. Prob. 17.)

Th. 3.2.1.9 Any homomorph of a cyclic group is cyclic. (Cf. Prob. 20.)

Many other interesting and important properties of cyclic groups will appear as problems (cf. Prob. 6 ff.) in the following pages.

Solved Problems

1. Prove Th. 3.2.1.1.

PROOF:

Since necessarily $a \in G$ if $a \in S$, it immediately follows that $a^i \in G$, $i = 0, 1, \dots$. But, then, since G is a finite group (as is not “stated otherwise”, cf. Prob. 7 below), the elements in the infinite sequence a^0, a^1, a^2, \dots cannot be all distinct. Hence it must be the case that $a^m = a^n$, where $m \neq n$; i.e. $a^{m-n} = a^{n-n} = a^0 = e$. Since $aa^{-1} = e = a^{m-n}$, it follows that $aa^{m-n-1} = e$. Now, let $m > n$; then $k = m - n - 1 \geq 0$. If $k > 0$, then $a^{-1} = a^k \in G$; and if $k = 0$, then $a^{-1} = a^0 = e \in G$. In either case it follows that $a^{-1} \in G$, satisfying the second condition of Th. 3.1.1.13. (The second condition, thus deducible from the first condition of Th. 3.1.1.13, is now proved to be redundant for finite groups.)

Thus satisfying both conditions of Th. 3.1.1.13, S is a subgroup of G if the product of any two elements of S is again in S .

Conversely, if S is a subgroup of G , then $a, b \in S$ obviously implies $ab \in S$, completing the proof for Th. 3.2.1.1.

Note that if $m < n$, the proof will be kept the same by a trivial revision: $k = n - m - 1$. Note, also, that this proof is valid iff G is a finite group; e.g. if G is the group of all integers under addition and S the set of all positive integers under addition, then both G and S satisfy the first condition, but not the second.

2. Prove Th. 3.2.1.2.

PROOF:

- (i) If S is a subgroup of G , then, by Th. 3.1.1.13, $a, b \in S$ implies $a^{-1} \in S$; hence $a^{-1}b \in S$.
- (ii) If $a, b \in S$ implies $a^{-1}b \in S$, then $a^{-1}a = e \in S$ which in turn implies $a^{-1}e = a^{-1} \in S$, which then implies $(a^{-1})^{-1}b = ab \in S$. Hence G1, G3, and G4 are established, and including G2, which obviously holds, the proof is complete.

3. If a group G is isomorphically mapped into a group G' , then a complex S' of G' , whose elements 1-1 correspond to the elements of a subgroup S of G , is a subgroup of G' .

PROOF:

Since $a, b \in S$ and $a', b' \in S'$ imply $a \leftrightarrow a'$ and $b \leftrightarrow b'$, and also, by Th. 3.1.3.9, $a^{-1} \in G$ and $a'^{-1} \in G'$ imply $a^{-1} \leftrightarrow a'^{-1}$, it follows that $a^{-1}b \leftrightarrow a'^{-1}b'$. But then, by Th. 3.2.1.2, $a^{-1}b \in S$; hence $a'^{-1}b' \in S'$ and, again by Th. 3.2.1.2, S' is a subgroup of G' .

4. Prove Th. 3.2.1.3.

PROOF:

By Df. 2.3.2, $a \in A \cap B$ implies $a \in A$ and $a \in B$, and consequently $a^{-1} \in A$ and $a^{-1} \in B$. Hence $aa^{-1} = e \in A$ and $e \in B$, which proves $A \cap B$ is not an empty set.

Furthermore, $a, b \in A \cap B$ likewise implies $ab \in A$ and $ab \in B$, which together imply $ab \in A \cap B$.

Hence, by Th. 3.2.1.1, $A \cap B$ is a subgroup of G .

5. Prove Th. 3.2.1.4.

PROOF:

G1: If $m, n \in I$ and $a^m, a^n \in C$, then, by Prob. 17 of §3.1.1, $a^m a^n = a^{m+n}$ and, since $m+n \in I$, it follows that $a^{m+n} \in C$, i.e. $a^m a^n \in C$.

G2: $a^l(a^m a^n) = (a^l a^m) a^n = a^{l+m} a^n = a^{l+m+n} \in C$

G3: $a^0 a^n = a^n a^0 = a^n$, and obviously $a^0 \in C$.

G4: $(a^n)^{-1} = a^{-n}$, by Prob. 18 of §3.1.1, and $a^{-n} \in C$.

Hence C forms a group.

6. Find the order of (i) an element -1 of the multiplicative group G whose elements are -1 and 1 , and also of (ii) an element 1 of the additive group I whose elements are all integers.

Solution:

- (i) Since $(-1)^2 = 1 = e \in G$, the order of -1 is 2.
 (ii) Define the additive operation of I as $(1)^n = n$; then, for no positive number n , $(1)^n = 0$ which is the identity of I . Hence the order of 1 is infinity (cf. Prob. 8 below), and the cyclic additive group generated by 1 is an infinite cyclic group.

In the following pages, as before, groups will be of finite order, unless stated otherwise.

7. A cyclic group generated by a is finite if $m \neq n$, $m, n \in I$, implies $a^m = a^n$, and infinite if $m \neq n$ implies $a^m \neq a^n$.

PROOF:

Since $m \neq n$, it is evidently the case that either $m > n$ or $m < n$. Let $m > n$ (or, with only a very slight modification, $m < n$, cf. Prob. 1 above); then $a^{m-n} = e$, $m-n > 0$, if $a^m = a^n$. Now, let k be the smallest of $m-n$ such that $a^k = e$, and p be any positive number such that $p = kq + r$, $r = 0, 1, \dots, k-1$; then, by Prob. 17, 18 of §3.1.1,

$$a^p = a^{kq+r} = a^{kq}a^r = (a^k)^qa^r = e^qa^r = ea^r = a^r$$

i.e. the positive integral exponents of the generator a will always be limited to $0, 1, \dots, k-1$. Hence the elements of C cannot but be: $a^0 = e, a, a^2, \dots, a^{k-1}$, which then of course forms a finite group.

If $m \neq n$ implies $a^m \neq a^n$, then, because of the proof just completed, C cannot be a finite group, i.e. must be an infinite group.

8. There exists an isomorphism between any infinite cyclic group C and the additive group I of all integers.

PROOF:

If $a^m, a^n \in C$ and $m, n \in I$ where $m \neq n$ and $a^m \neq a^n$ (cf. Prob. 7 above), i.e. a^m and a^n are distinct elements of C , then let $a^m \leftrightarrow m$ and $a^n \leftrightarrow n$, which in turn imply $a^m a^n = a^{m+n} \in C \leftrightarrow m+n \in I$, completing the proof.

9. Any infinite cyclic group has exactly two distinct generators: one generator and its inverse.

PROOF:

Let the one generator be a ; then, since $a^n = (a^{-1})^{-n}$ by Prob. 14 of §3.1.1, a^{-1} is obviously the other generator.

They are distinct, i.e. $a \neq a^{-1}$, since $a = a^{-1}$ implies $aa^{-1} = a^2 = e$ (i.e. a finite cyclic group of order 2), which is a contradiction.

They also defy the existence of any other possible generators for the following reason. Suppose there exists the third generator b of the same group; then $a = b^i$ and $b = a^j$ (since both a and b are generators), i.e.,

$$a = (a^j)^i = a^{ji}$$

but, by Prob. 7 above (" $m \neq n$ implies $a^m \neq a^n$ " is, by the contrapositive rule, cf. MTh.1.1.1.12, logically equivalent to " $a^m = a^n$ implies $m = n$ "), it must be the case that $1 = ji$, where i and j are both integers.

Hence j must be either 1 or -1 , i.e. $b = a$ or $b = a^{-1}$, proving that a and a^{-1} exhaust all possible generators.

10. The additive group R^* of all real numbers is not cyclic; nor is the octic group D_4 of the square (cf. Prob. 5 of §3.1.2).

PROOF:

- (i) Suppose R^* be cyclic, and let $a \neq 0 \in R^*$ be a generator; then, if the additive operation of R^* is defined as $(a)^n = na$, where $n \in I$ (cf. Prob. 6 above), it follows that $|(a)^n| \geq |a|$, i.e. any real number smaller than $|a|$ cannot belong to R^* , which is a contradiction. Hence the additive group R^* in its entirety cannot be cyclic.

- (ii) Of the right transformations with respect to rotational and reflective symmetries: 0, 1, 2, 3, 4, 5, 6, 7, none can generate the other seven; hence, by definition, it is not cyclic.

(Note, however, that the group can be generated by two: 1 and 5; viz. $0 = 1^0$, $1 = 1$, $2 = 1^2$, $3 = 1^3$, $4 = 51$, $5 = 5$, $6 = 51^2$, $7 = 51^3$.)

11. Verify that the multiplication table on the right specifies a cyclic group, by finding its generator (or generators).

PROOF:

(i) $a = a$, $a^2 = c$, $a^3 = d$, $a^4 = a^0 = b$.

(ii) $d = d$, $d^2 = c$, $d^3 = a$, $d^4 = d^0 = b$.

	a	b	c	d
a	c	a	d	b
b	a	b	c	d
c	d	c	b	a
d	b	d	a	c

12. If a is a generator of a cyclic group C of order k , then $a, a^2, \dots, a^k = e$ are all distinct elements of C ; and if in general $a^p = e$, then p is divisible by k .

PROOF:

- (i) If the elements of C are not distinct, i.e. if some, e.g. a^i and a^j , are identical, then $a^{i-j} = e$, which implies that $i-j$ (if $i > j$, or what is the same: $j-i$ if $i < j$) is smaller than k , i.e. that k is not the smallest positive integer for which $a^k = e$, contradictory to Df. 3.2.1.6. Hence the elements of C must be all distinct.

- (ii) Since $a^p = a^r = e$ for any positive number p such that $p = kq + r$ (cf. Prob. 7 above), r must be now either smaller than k , which is against Df. 3.2.1.6, or 0. Hence r must be 0, i.e. $p = kq$, which is evidently divisible by k .

13. Let the group R of all rotations of a regular octagon through $\pi/4$ be $R_0, R_1, R_2, R_3, R_4, R_5, R_6, R_7$, and find the elements among them which generate R .

Solution:

Since $R_1^8 = R_0 = (1)$, R_1 is a generator which does generate the other seven distinct elements; but R_2, R_4, R_6 are not, since $R_2^4 = R_4^2 = R_6^4 = (1)$, against the conditions prescribed by Prob. 12 above. R_3 , however, generates distinct elements, viz. $R_3^8 = R_0$, $R_3^3 = R_1$, $R_3^6 = R_2$, $R_3^5 = R_3$, $R_3^4 = R_4$, $R_3^7 = R_5$, $R_3^2 = R_6$, $R_3^5 = R_7$.

Likewise, R_5 and R_7 generate C .

Note, as will be generalized below, that the greatest common divisor of the exponents of the generators, i.e. 1, 3, 5, 7, and the order of R , namely 8, is 1.

14. If a cyclic group C is generated by an element a of order n , then a^m generates C iff the greatest common divisor of m and n is 1.

PROOF:

The n elements, $a^m, a^{2m}, \dots, a^{(n-1)m}, a^{nm} = e$, are all distinct. For, if e.g. $a^{im} = a^{jm}$, $0 \leq i, j \leq n$ and $i > j$ (cf. Prob. 1, 7, above), then

$$a^{im-jm} = a^{(i-j)m} = e$$

where, by Prob. 12, $(i-j)m$ must be divisible by n and, by hypothesis (viz. the g.c.d. of m and n is 1), $i-j$ itself must be divisible by n . But then, by the original stipulation, $i-j$ is a positive integer smaller than n , which brings forth a contradiction. Hence the n elements $a^m, a^{2m}, \dots, a^{(n-1)m}, e$ are all distinct; i.e. a^m is a generator of C .

Conversely, if a^m is a generator of C , let the g.c.d. of m and n be d and $d \neq 1$, i.e. $d > 1$. Then, since m/d and n/d must be positive integers and also, by Prob. 12, $a^n = e$,

$$(a^m)^{n/d} = (a^n)^{m/d} = e^{m/d} = e$$

where n/d is necessarily a positive integer smaller than n itself. But then, again by Prob. 12, a^m cannot be a generator of C , contrary to the assumption. Hence d must be 1, completing the proof.

15. Any group G of order 3 is cyclic.

PROOF:

Since G is of order 3, it must have elements other than e . Let $a \neq e \in G$, then $a^2 \neq a$; for, otherwise, it must be the case that $a = e$ (cf. §3.1.1, Prob. 7), contrary to the assumption.

Also $a^2 \neq e$; for, otherwise (i.e. if $a^2 = e$), it must be the case, by definition (of a group), that $ab = a$ or $ab = b$ or $ab = e$ for the third element b of G , i.e. $b \neq e$ and $b \neq a$. But if $ab = a$, then $b = e$, and if $ab = b$, then $a = e$, both contrary to the assumption. Hence, eliminating the two cases out of three, it must be the case that $ab = e$. But then, if $a^2 = e$, it follows that $a^2 = ab$, i.e. $a = b$, again contrary to the assumption. Hence $a^2 \neq a$, and e, a, a^2 are three distinct elements of G .

Since, then, $aa^2 = a^3 = e$, it follows that a must be the generator of G ; hence G is cyclic, completing the proof.

16. Any cyclic group is Abelian.

PROOF:

If a is a generator of a cyclic group C , i.e. $a^m, a^n \in C$, for any $m, n \in I$, then

$$a^m a^n = a^{m+n} = a^{n+m} = a^n a^m$$

Hence C is Abelian.

17. Any subgroup of a cyclic group is also cyclic.

PROOF:

Let S be a subgroup of a cyclic group C ; then $a \in C$ implies that the elements of S are all given in terms of a^p , where $p > 0$. Given in general $p = kq + r$, $0 \leq r < k$, p must always be divisible by k , i.e. $r = 0$, if k is to be the smallest of p . Hence $a^p = (a^k)^q$, which proves a^k to be a generator of S . S itself, then, is a cyclic group.

Note, as in Prob. 1, 7, 14, etc., that the case of $p < 0$ does not affect the proof as a whole; for, then, $a^{-p} \in S$, $-p > 0$, and $a^{-p} = (a^k)^q$, i.e. $a^p = (a^k)^{-q} \in S$.

18. Set up a homomorphism between the alternating group A_4 (cf. Df. 3.1.2.18) and the cyclic group of order 3.

Solution:

Given A_4 : $(1), (123), (132), (124), (142), (134), (143), (234), (243), (12)(34), (13)(24), (14)(23)$, it is at once observed that $(123)^2 = (132)$, $(124)^2 = (142)$, $(134)^2 = (143)$, $(234)^2 = (243)$, and also that $((12)(34))^3 = (12)(34)$, $((13)(24))^3 = (13)(24)$, $((14)(23))^3 = (14)(23)$, and of course $(1)^3 = (1)$.

Hence, the cyclic group of order 3 being the set $\{a, a^2, a^3 = e\}$, there exists a homomorphism of three four-to-one correspondences:

$$\begin{aligned} (1), (12)(34), (13)(24), (14)(23) &\rightarrow e = a^3 \\ (123), (124), (134), (234) &\rightarrow a \\ (123)^2, (124)^2, (134)^2, (234)^2 &\rightarrow a^2 \end{aligned}$$

19. Verify the homomorphism of the cyclic group of order 2 onto the dihedral group D_4 (i.e. the octic group of the square).

PROOF:

Referring to §3.1.2, Prob. 5, 6, rewrite:

$$e = (1), \quad a = (1234), \quad b = (13)(24), \quad c = (1432), \quad d = (12)(34), \quad f = (14)(23), \quad g = (13), \quad h = (24)$$

which in turn are reduced to

$$e = a^4, \quad a = a, \quad b = a^2, \quad c = a^3, \quad d = a^2f, \quad f^2 = e, \quad g = a^3f, \quad h = af$$

Hence there exists a homomorphism of two four-to-one correspondences, viz.,

$$a, a^2, a^3, a^4 \rightarrow e \quad \text{and} \quad af, a^2f, a^3f, a^4f \rightarrow a$$

20. Prove Th. 3.2.1.9.

PROOF:

Since, by Th. 3.1.3.1 and Th. 3.1.3.3, any homomorphism $H: G \rightarrow G'$ maps the identity and the inverses of G onto their counterparts of G' , it is always the case that $(a^n)' = (a')^n$, $n \in I$, under H if a is the generator of G . Hence the powers $(a')^n = (a^n)'$ of a' exhaust G' as the powers a^n exhaust G .

§3.2.2 Cosets and Conjugates

Df. 3.2.2.1 The *product* of two complexes K_1 and K_2 , denoted by K_1K_2 , of a group G is the set K of all elements of the form k_ik_j , where $k_i \in K_1$, $k_j \in K_2$, and k_ik_j may not be all distinct.

If K_1 and K_2 are both finite, i.e. $i = 1, 2, \dots, m$ and $j = 1, 2, \dots, n$, then K has mn elements, not all of them distinct; e.g. if G is the symmetric group S_3 of degree 3, $K_1 = \{(1), (12), (123)\}$, and $K_2 = \{(13), (132)\}$, then K has $2 \cdot 3 = 6$ elements, of which only 4 elements are distinct, viz. $(1)(13) = (12)(132) = (13)$, $(1)(132) = (12)(13) = (132)$, $(123)(13) = (23)$, $(123)(132) = (1)$.

Complexes are special subsets, since they are specified as “non-empty” subsets (cf. Df. 3.1.1.11), and as such need a slight modification, as has already been observed in Df. 3.2.2.1, with respect to their joins and meets. The products of complexes, in fact, are no longer the meets as defined in Df. 2.3.2, but the sums of complexes are still the joins as defined in Df. 2.3.1; viz. for every $K_1, K_2, K_3 \subset G$,

$$K_1 \cup K_1 = K_1, \quad K_1 \cup K_2 = K_2 \cup K_1, \quad K_1 \cup (K_2 \cup K_3) = (K_1 \cup K_2) \cup K_3$$

Multiplication of complexes is generally non-commutative, but definitely associative and distributive, as is plainly justified by Df. 3.2.2.1, viz.,

$$K_1(K_2K_3) = (K_1K_2)K_3, \quad K_1(K_2 \cup K_3) = K_1K_2 \cup K_1K_3, \quad (K_2 \cup K_3)K_1 = K_2K_1 \cup K_3K_1$$

The inclusion relation also holds; viz. if each element of K_1 is an element of K_2 , then $K_1 \subseteq K_2$, and if the converse $K_2 \subseteq K_1$ simultaneously holds, then $K_1 = K_2$.

The cancellation law does not hold here; i.e. $K_1K_2 = K_1K_3$ or $K_2K_1 = K_3K_1$ does not imply $K_2 = K_3$, although, conversely, $K_2 = K_3$ implies $K_1K_2 = K_1K_3$ and $K_2K_1 = K_3K_1$ iff K_1 consists of a single element.

Df. 3.2.2.2 If G_1 is a subgroup of G and K_1 a single-element complex of G , then K_1G_1 is called a *left-coset*, and G_1K_1 a *right-coset*, of G_1 in G .

G_1 itself is both a left- and right-coset of G_1 in G , since $eG_1 = G_1e = G_1$ for $K_1 = \{e\}$; also, if K_1 is any single-element complex of G , both K_1G_1 and G_1K_1 must contain K_1 itself, since $e \in G_1$.

Note. “Right-coset” and “left-coset” are duals, as is implied by Df. 3.2.2.2 which, together with Df. 3.1.1.7, also implies that such an orientation of right and left is needed only for non-commutative groups; for, in an Abelian group, right-cosets and left-cosets cannot but coincide. Any of the following theorems stated in terms of left-cosets will thus have a counterpart in terms of right-cosets.

In the following theorems the single-element complexes K_i will be replaced by single letters a, b, c, \dots .

Th. 3.2.2.3 There exists a 1-1 correspondence between a subgroup G_1 of G and a left-coset aG_1 in G (cf. Prob. 1).

Th. 3.2.2.4 Iff $a^{-1}b \in G_1$, then $aG_1 = bG_1$ (cf. Prob. 2).

As a matter of fact, $a^{-1}b \in G_1$ implies $a \equiv b$ (cf. Prob. 3).

Th. 3.2.2.5 If $aG_1 \neq bG_1$, then aG_1 and bG_1 are disjoint (cf. Prob. 4).

Th. 3.2.2.4 and Th. 3.2.2.5, when put together, assert: any two left-cosets aG_1 and bG_1 are either identical or without common elements.

Th. 3.2.2.6 There exists a partition (cf. Df. 2.3.13) of G ,

$$G = a_1 G_1 \cup a_2 G_1 \cup \dots \cup a_n G_1$$

where G is decomposed into a join of a finite number of disjoint cosets (cf. Prob. 6-7).

Df. 3.2.2.7 The partition in Th. 3.2.2.6 is called the *left-decomposition* of G with respect to G_1 , and an arbitrary element a_i of each left-coset is called a *representative* (cf. Df. 2.1.15) of that coset. The equation of the partition itself is called a *class equation*.

The orientation of right and left is necessary for the decomposition of non-commutative groups, because right and left decompositions may turn out to be distinct. (Cf. Prob. 7, 10, 11.)

Th. 3.2.2.8 If, as in Th. 3.2.2.6, $G = a_1 G_1 \cup a_2 G_1 \cup \dots \cup a_n G_1$, then

$$G = G_1 a_1^{-1} \cup G_1 a_2^{-1} \cup \dots \cup G_1 a_n^{-1}$$

Df. 3.2.2.9 The number n of distinct left-cosets of G_1 in G is called the *index* of G_1 in G , denoted by $n = (G : G_1)$.

Example:

The index of the left-cosets of G_1 in G above, in Th. 3.2.2.8, equals the index of the right-cosets of G_1 (cf. Prob. 9). In the case of infinite groups, however, the sets of right and left cosets with respect to a subgroup are said to have the same cardinal number.

Th. 3.2.2.10 (by Lagrange). The order of a subgroup G_1 of a group G is a divisor of the order of G (cf. Prob. 12-13).

Conversely, however, a group of order n may not have a subgroup of order k even if k is a divisor of n ; e.g. a permutation group of order 12,

$$(1), (123), (124), (132), (142), (234), (243), (134), (143), (12)(34), (13)(24), (14)(23)$$

has no subgroup of order 6, although it does have subgroups of orders 2, 3, 4. (The converse holds if k is a prime or prime power; the existence and number of such subgroups is studied by the Sylow theorems, which are beyond the scope of this book.)

Th. 3.2.2.11 The order of an element $a \in G$ is a factor of the order of G . (Cf. Prob. 14.)

Example:

In the octic group of the square (cf. §3.1.2, Prob. 5) the transformation 1 has order 4 and other transformations also have orders (such as 2) which are factors of 8.

Th. 3.2.2.12 Any group of prime order is cyclic and has no proper subgroups. (Cf. Prob. 15.)

* * * * *

Df. 3.2.2.13 If K is a complex of a group G , then the set H whose elements are of the form $x^{-1}kx$, for some x and all $k \in K$, is called the *transform* of K by x , denoted by $H = x^{-1}Kx$ or $H = K^x$; the element x , which need not be unique, is called the *transforming element*.

Df. 3.2.2.14 If both H and K are complexes of G in Df. 3.2.2.13, and if G_1 is a subgroup of G such that $x \in G_1$, then H is said to be the *conjugate* of K under G_1 ; all H conjugate to K under G_1 constitutes a *class of conjugates*.

Example:

If G is S_3 , $G_1 = \{(1), (12)\}$, and $K = \{(12), (13), (23)\}$, then two sets of conjugates of K under G_1 : $H_1 = (1)^{-1}\{(12), (13), (23)\}(1)$ and $H_2 = (12)^{-1}\{(12), (13), (23)\}(12)$ yield a class of conjugates.

If $a, b, c \in G$, then the conjugate c of b by a under G is $c = a^{-1}ba$.

Example:

In S_3 , (23) is the conjugate of (12) by (123) , since $(23) = (123)^{-1}(12)(123)$.

In general, if

$$p = \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ b_1 & b_2 & \dots & b_n \end{pmatrix} \quad \text{and} \quad q = \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ c_1 & c_2 & \dots & c_n \end{pmatrix} = \begin{pmatrix} b_1 & b_2 & \dots & b_n \\ d_1 & d_2 & \dots & d_n \end{pmatrix}, \quad \text{then} \quad q^{-1} = \begin{pmatrix} c_1 & c_2 & \dots & c_n \\ a_1 & a_2 & \dots & a_n \end{pmatrix}$$

(cf. Th. 3.1.2.6)

$$\text{and} \quad r = q^{-1}pq = \begin{pmatrix} c_1 & c_2 & \dots & c_n \\ a_1 & a_2 & \dots & a_n \end{pmatrix} \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ b_1 & b_2 & \dots & b_n \end{pmatrix} \begin{pmatrix} b_1 & b_2 & \dots & b_n \\ d_1 & d_2 & \dots & d_n \end{pmatrix} = \begin{pmatrix} c_1 & c_2 & \dots & c_n \\ d_1 & d_2 & \dots & d_n \end{pmatrix}$$

Th. 3.2.2.15 Conjugacy is an equivalence relation. (Cf. Prob. 16.)

Th. 3.2.2.16 K and K^x in Df. 3.2.2.13 are of the same order. (Cf. Prob. 17.)

Th. 3.2.2.17 If S is the set of all cycles of length n , then $S = S^x$ (cf. Prob. 18).

Th. 3.2.2.18 The order of an element $c \in G$ which is conjugate to an element $a \in G$ equals the order of a itself. (Cf. Prob. 21.)

Solved Problems

1. Prove Th. 3.2.2.3.

PROOF:

Let $g_i \in G_1$ and $ag_i \in aG_1$; then, since $ag_i = ag_j$ at once implies $g_i = g_j$, the correspondence $g_i \leftrightarrow ag_i$ must be 1-1.

Stated otherwise: each element $a_i = ag_i$ of the coset aG_1 is in fact the unique image of a distinct element g_i of G_1 , viz. $g_i = a^{-1}a_i$. Hence $g_i \leftrightarrow a_i$, or what is the same, $a^{-1}a_i \leftrightarrow ag_i$.

2. Prove Th. 3.2.2.4.

PROOF:

If $a^{-1}b \in G_1$, then $(a^{-1}b)^{-1} \in G_1$. Let $a^{-1}b = g_1$ as in general $g_i \in G_1$, $i = 1, 2, \dots, n$; then $(a^{-1}b)^{-1} = b^{-1}(a^{-1})^{-1} = b^{-1}a$ (cf. §3.1.1, Prob. 19), i.e. $b^{-1}a = (a^{-1}b)^{-1} = g_1^{-1} \in G_1$. Also, if $a_j \in aG_1$, $j = 1, 2, \dots, n$, then

$$a_j = ag_i = (bg_1^{-1})ag_i = b(b^{-1}a)g_i = bg_1^{-1}g_j \in bG_1, \quad \text{i.e.} \quad a_j \in aG_1 \rightarrow a_j \in bG_1$$

Hence $aG_1 \subseteq bG_1$. Likewise, $bG_1 \subseteq aG_1$, starting with $b_j \in bG_1$ to arrive at $b_j \in aG_1$. Hence $aG_1 = bG_1$.

Conversely, if $aG_1 = bG_1$ to begin with, then $g_i \in G_1$ and $ag_i = bg_i$, which implies $a^{-1}ag_i = a^{-1}bg_i$, which in turn implies $g_i = a^{-1}bg_i$, which further implies $a^{-1}b = g_i g_i^{-1} \in G_1$. Hence $aG_1 = bG_1$ implies $a^{-1}b \in G_1$.

3. Examine $a^{-1}b \in G_1$ in terms of an equivalence relation.

Solution:

- (i) Reflexivity: $a \equiv a$ implies $a^{-1}a = e \in G_1$.
- (ii) Symmetry: $a \equiv b$ implies $a^{-1}b \in G_1$, which in turn implies $b^{-1}a = (a^{-1}b)^{-1} \in G_1$, which further implies $b \equiv a$.
- (iii) Transitivity: $a \equiv b$ and $b \equiv c$ imply $a^{-1}b \in G_1$ and $b^{-1}c \in G_1$, which in turn imply $a^{-1}c = (a^{-1}b)(b^{-1}c) \in G_1$, which further implies $a \equiv c$.

Furthermore, since the implications in (i)-(iii) can be readily reversed, the converses of (i)-(iii) also hold. Hence there exists not only an equivalence relation with respect to $a^{-1}b$, but also the latter implies $a \equiv b$ in G_1 .

4. Prove Th. 3.2.2.5.

PROOF:

Suppose $aG_1 \cap bG_1 \neq \emptyset$, i.e. suppose that aG_1 and bG_1 do have an element x in common; then $x = ag_p = bg_q$ for some p and q where $ag_p, bg_q \in G_1$. Hence $a^{-1}ag_p = a^{-1}bg_q$, i.e. $a^{-1}b = g_p g_q^{-1} \in G_1$, and, by Th. 3.2.2.4, $aG_1 = bG_1$, contrary to the initial assumption. Hence it must be the case that $aG_1 \cap bG_1 = \emptyset$.

5. A subgroup G_1 of a group G is the only coset which in itself is a subgroup of G .**PROOF:**

Let any coset of G_1 be aG_1 ; then, if aG_1 is a subgroup, $e \in G_1$ implies $e \in aG_1$ (cf. §3.1.1, Prob. 8), i.e. $G_1 \subseteq aG_1$, and conversely, i.e. $aG_1 \subseteq G_1$. Hence $aG_1 = G_1$.

6. Prove Th. 3.2.2.6.

PROOF:

Let the elements of G be b_1, b_2, \dots, b_r , $r \geq n$; then there follow r left-cosets $b_1G_1, b_2G_1, \dots, b_rG_1$, which may not be mutually exclusive (cf. Th. 3.2.2.5-6). Hence choose out of r left-cosets only n non-overlapping left-cosets a_iG_1 , $i = 1, 2, \dots, n$, and

$$G = \cup_{i \in I} a_iG_1, \quad I = \{1, 2, \dots, n\}$$

7. Find the left decomposition of the octic group D_4 of the square (cf. §3.1.2, Prob. 5, 6): $e = (1)$, $a = (1234)$, $b = (13)(24)$, $c = (1432)$, $d = (12)(34)$, $f = (14)(23)$, $g = (13)$, $h = (24)$, with respect to a subgroup $G_1 = \{e, h\}$.**Solution:**

Since there are eight elements in D_4 : e, a, b, c, d, f, g, h (corresponding to b_1, b_2, \dots, b_r in Prob. 6 above), there follow eight left cosets (corresponding to $b_1G_1, b_2G_1, \dots, b_rG_1$), viz.,

$$\begin{aligned} eG_1 &= \{ee, eh\} = \{e, h\} & aG_1 &= \{ae, ah\} = \{a, f\} & bG_1 &= \{be, bh\} = \{b, g\} \\ cG_1 &= \{ce, ch\} = \{c, d\} & dG_1 &= \{de, dh\} = \{d, c\} & fG_1 &= \{fe, fh\} = \{f, a\} \\ gG_1 &= \{ge, gh\} = \{g, b\} & hG_1 &= \{he, hh\} = \{h, e\}. \end{aligned}$$

The eight left-cosets, however, are not mutually exclusive, as can be readily observed, since $eG_1 = hG_1$, $aG_1 = fG_1$, $bG_1 = gG_1$, $cG_1 = dG_1$. Hence choose only 4 disjoint left-cosets (corresponding to a_iG_1 , $i = 1, 2, 3, 4$) out of eight, viz. either the set of eG_1, aG_1, bG_1, cG_1 , or the set of hG_1, fG_1, gG_1, dG_1 . In either case the partition of D_4 is complete, viz.,

$$D_4 = eG_1 \cup aG_1 \cup bG_1 \cup cG_1 = dG_1 \cup fG_1 \cup gG_1 \cup hG_1$$

8. Prove Th. 3.2.2.8.

PROOF:

If $b \in G_1$, then $b^{-1} \in a_iG_1$, $i = 1, 2, \dots, n$, which implies that there exists $g \in G_1$ such that $b^{-1} = a_i g$. Since $b = (b^{-1})^{-1} = (a_i g)^{-1} = g^{-1} a_i^{-1}$, where $g^{-1} \in G_1$, it follows that $b \in G_1 a_i^{-1}$. Hence

$$G = G_1 a_1^{-1} \cup G_1 a_2^{-1} \cup \dots \cup G_1 a_n^{-1}$$

where all right-cosets are mutually exclusive. For, if $G_1 a_i^{-1} \cap G_1 a_j^{-1} \neq \emptyset$ and there thus exists $c \in G$ such that $c \in G_1 a_i^{-1}$ and $c \in G_1 a_j^{-1}$, then $c = g_1 a_i^{-1}$ and $c = g_2 a_j^{-1}$ for some $g_1, g_2 \in G_1$; hence

$$a_i g_1^{-1} = (g_1 a_i^{-1})^{-1} = (g_2 a_j^{-1})^{-1} = a_j g_2^{-1}$$

which implies $a_i G_1 \cap a_j G_1 \neq \emptyset$, contrary to the original hypothesis. Hence the n right-cosets must be all mutually exclusive, and there now exists the right decomposition of G with respect to G_1 , where the number of right-cosets is equal to that of left-cosets.

9. There exists a 1-1 correspondence between the right and left cosets of G_1 in G .**PROOF:**

It immediately follows, from Prob. 8 above, that there exists a 1-1 correspondence of the form $a_i G_1 \leftrightarrow G_1 a_i^{-1}$.

10. Decompose the octic group of the square in Prob. 6 into a join of right-cosets.

Solution:

Since $e^{-1} = e$, $a^{-1} = c$, $b^{-1} = b$, $c^{-1} = a$, $d^{-1} = d$, $f^{-1} = f$, $g^{-1} = g$, $h^{-1} = h$, it follows that

$$G = eG_1 \cup aG_1 \cup bG_1 \cup cG_1 = G_1e^{-1} \cup G_1a^{-1} \cup G_1b^{-1} \cup G_1c^{-1} = G_1e \cup G_1c \cup G_1b \cup G_1a$$

and also, likewise, $G = dG_1 \cup fG_1 \cup gG_1 \cup hG_1 = G_1d \cup G_1f \cup G_1g \cup G_1h$.

11. Find the difference, if any, between the right and left decompositions of the symmetric group G of degree 3, viz. $(1), (12), (13), (23), (123), (132)$, with respect to a subgroup $G_1 = \{(1), (12)\}$.

Solution:

According to the left decomposition:

$$G = (1)G_1 \cup (13)G_1 \cup (23)G_1 = \{(1), (12)\} \cup \{(13), (132)\} \cup \{(23), (123)\}$$

and according to the right decomposition:

$$G = G_1(1) \cup G_1(13) \cup G_1(23) = \{(1), (12)\} \cup \{(13), (123)\} \cup \{(23), (132)\}$$

Hence the two decompositions are distinct, although they do consist of the same number of cosets.

12. The order of a group G equals the product of the index of a subgroup G_1 in G and the order of G_1 .

PROOF:

Since there always exists a partition of G into n disjoint cosets (cf. Prob. 6-7), each of which in turn has exactly m elements, m being the order of G_1 itself (cf. Th. 3.2.2.3), mn must be the order of G .

13. Prove Th. 3.2.2.10.

PROOF:

Since $\{e\}$, the set which consists of e alone, is a subgroup of any group, the orders of G and G_1 may be written in terms of indices, viz. $(G:\{e\})$ and $(G_1:\{e\})$ respectively (cf. Df. 3.2.2.9, and also note that any coset aG_1 contains $a = ae$). Then, by Prob. 12 above,

$$(G:\{e\}) = (G_1:\{e\})(G:G_1)$$

Hence the index $(G:G_1)$ must be a divisor of the index $(G:\{e\})$.

14. Prove Th. 3.2.2.11.

PROOF:

If $a \in G$, then a cyclic subgroup generated by a has the same order as the order of a itself (cf. Df. 3.2.1.6). Hence, by Th. 3.2.2.10, the order of a divides the order of G .

15. Prove Th. 3.2.2.12.

PROOF:

Since, by Th. 3.2.2.11, the order of a cyclic subgroup G_1 generated by an element $a \in G$ ($a \neq e$) must be a factor of the prime order p of G , when p has as its factors only 1 and p itself, the order of G_1 must be p . (Note. $a \neq e$, i.e. G_1 has a as well as e , and as such cannot be of order 1.) Hence, the orders of G and G_1 being the same, $G = G_1$, and G must then be cyclic.

If G has a proper subgroup G_2 , then the order k of G_2 must be a factor of the order p of G . But then, since k can be neither p nor 1 (cf. Df. 3.1.1.12), it is self-contradictory. Hence G cannot have any proper subgroup.

16. Prove Th. 3.2.2.15.

PROOF:

- (i) Reflexivity: $a_1 = e^{-1}a_1e$, where $a_1 \in G$; i.e. a_1 is the conjugate of itself by e .
- (ii) Symmetry: If, for $a_1, a_2 \in G$ and for some b , $a_1 = b^{-1}a_2b$, then $(b^{-1})^{-1}a_1b^{-1} = (b^{-1})^{-1}(b^{-1}a_2b)b^{-1} = a_2$; i.e. if a_1 is the conjugate of a_2 by b , so is a_2 of a_1 by b^{-1} .
- (iii) Transitivity: If, for $a_1, a_2, a_3 \in G$ and for some b, c , $a_1 = b^{-1}a_2b$ and $a_2 = c^{-1}a_3c$, then $a_1 = b^{-1}(c^{-1}a_3c)b = (cb)^{-1}a_3(cb)$; i.e. if a_1 is the conjugate of a_2 by b , and a_2 the conjugate of a_3 by c , then a_1 is the conjugate of a_3 by cb .

17. Prove Th. 3.2.2.16.

PROOF:

Since there exists a correspondence, for any $s \in S, r \in R$, and some $x \in S$, such that $s \rightarrow r = x^{-1}sx$, and conversely that $r \rightarrow s = r x r^{-1} = x(x^{-1}sx)x^{-1} = s$, the correspondence between S and $S^x = R$ is one-one. Hence both sets have the same number of elements.

18. Prove Th. 3.2.2.17.

PROOF:

Let $s = (a_1 a_2 \dots a_n)$ and $x = \begin{pmatrix} a_1 & a_2 & \dots & a_n & \dots & a_r \\ b_1 & b_2 & \dots & b_n & \dots & b_r \end{pmatrix}$; then

$$\begin{aligned} x^{-1}sx &= \begin{pmatrix} b_1 & b_2 & \dots & b_n & \dots & b_r \\ a_1 & a_2 & \dots & a_n & \dots & a_r \end{pmatrix} \begin{pmatrix} a_1 & a_2 & \dots & a_n & \dots & a_r \\ a_2 & a_3 & \dots & a_1 & \dots & a_r \end{pmatrix} \begin{pmatrix} a_1 & a_2 & \dots & a_n & \dots & a_r \\ b_1 & b_2 & \dots & b_n & \dots & b_r \end{pmatrix} \\ &= \begin{pmatrix} b_1 & b_2 & \dots & b_n & \dots & b_r \\ b_2 & b_3 & \dots & b_1 & \dots & b_r \end{pmatrix} = (b_1 b_2 \dots b_n) \end{aligned}$$

Hence any element of S^x is also a cycle of length n .

Conversely, it is clear that there exists some x such that $x^{-1}sx = r$ for any $r \in S^x$.

Hence $S = S^x$.

19. If a set S consists of the permutations of the form: $s = a_1 a_2 \dots a_n$ where a_1, a_2, \dots, a_n are disjoint cycles of lengths o_1, o_2, \dots, o_n respectively, then the set of all permutations conjugate to S contains the elements of the form: $r = b_1 b_2 \dots b_n$ where b_1, b_2, \dots, b_n are disjoint cycles of lengths o_1, o_2, \dots, o_n respectively.

PROOF:

For some c , $r = c^{-1}sc = c^{-1}(a_1 a_2 \dots a_n)c = c^{-1}(a_1(cc^{-1})a_2(cc^{-1}) \dots (cc^{-1})a_n)c = (c^{-1}a_1c)(c^{-1}a_2c) \dots (c^{-1}a_nc)$, where all conjugates of a_i , $i = 1, 2, \dots, n$, by c are disjoint since the set of a_i consists of disjoint cycles. Also, by Th. 2.2.2.17, each of $c^{-1}a_i c$ must have the length of a_i itself, and conversely, by Th. 3.2.2.16-17. Hence there exists a 1-1 correspondence between the two sets of disjoint cycles of the same lengths: $a_1 \leftrightarrow b_1, a_2 \leftrightarrow b_2, \dots, a_n \leftrightarrow b_n$.

20. Classify the classes of conjugates in the symmetric groups of degree 3, 4, 5.

Solution:

- (i) S_3 has six elements: $(1), (12), (13), (23), (123), (132)$, the conjugates of which are subdivided into three sets:
 - (a) (1) . This is trivial, since it always holds that $x^{-1}(1)x = (1)$.
 - (b) $(12), (13), (23)$. E.g. $(1)^{-1}(12)(1) = (12)$; $(23)^{-1}(12)(23) = (23)$, $(132)^{-1}(12)(132) = (13)$; i.e. the conjugates of three transpositions of S_3 are also transpositions, as can be deduced from Th. 3.2.2.17 and Prob. 19.
 - (c) $(123), (132)$. E.g. $(1)^{-1}(123)(1) = (123)$; $(23)^{-1}(123)(23) = (132)$; $(132)^{-1}(123)(132) = (123)$, which is indeed the expected result; i.e. the conjugates of the cycles in S_3 of length 3 are also the cycles of length 3.
- (ii) The conjugates of S_4 , likewise, are subdivided into five types: (a) the identity permutation (1) ; (b) the transpositions, e.g. (12) ; (c) the cycles of length 3, e.g. (123) ; (d) the cycles of length 4, e.g. (1234) ; (e) the product of two disjoint transpositions, e.g. $(12)(34)$.

- (iii) The conjugates of S_3 , likewise, are classified into seven patterns: $(a), (b), (c), (d), (e)$ as in (ii), and (f) the cycles of length 5, e.g. (12345) , (g) the products of one transposition and one cycle of length 3 which are disjoint, e.g. $(12)(345)$.

21. Prove Th. 3.2.2.18.

PROOF:

If $c = b^{-1}ab$ for some b and $a^n = e$, then, since for $n = 2$: $(b^{-1}ab)^2 = (b^{-1}ab)(b^{-1}ab) = b^{-1}a^2b$, suppose for $n = k$: $(b^{-1}ab)^k = b^{-1}a^kb$. Then, for $n = k + 1$, $(b^{-1}ab)^{k+1} = (b^{-1}ab)^k(b^{-1}ab) = (b^{-1}a^kb)(b^{-1}ab) = b^{-1}a^{k+1}b$, and in general, $(b^{-1}ab)^n = b^{-1}a^nb = b^{-1}eb = e$. Hence the order of $c = b^{-1}ab$ cannot be greater than that of a .

Conversely, if $(b^{-1}ab)^n = e$, then $b^{-1}a^nb = e$.

Hence $a^n = (bcb^{-1})^n = bc^nb^{-1} = b(b^{-1}ab)^nb^{-1} = b(b^{-1}a^nb)b^{-1} = beb^{-1} = e$, and the order of a cannot be greater than that of c .

Thus a and c must have the same order.

*§3.2.3 Normalizers and Centralizers

Th. 3.2.3.1 Given a subgroup G_1 of a group G , the complex K of all $x \in G$ such that $G_1^x = G_1$ is a subgroup of G . (Cf. Prob. 1.)

Since $G_1^x = G_1$, i.e. $x^{-1}G_1x = G_1$, is logically equivalent to $G_1x = xG_1$, Th. 3.2.3.1 has an alternative form: a complex K , $K \subseteq G$, which consists of those elements of G that commute with every element of G_1 is a subgroup of G .

Df. 3.2.3.2 The complex K of Th. 3.2.3.1 is called the *normalizer* of G_1 in G , denoted by $N_{G_1}(K)$ or simply $N(K)$.

Example:

If G is the symmetric group $S_3 = \{(1), (12), (13), (23), (132), (123)\}$ and $G_1 = \{(123), (132)\}$, then the normalizer $N_{G_1}(K)$ is the complex $K = \{(1), (123), (132)\}$, since $(1)(123) = (123)(1)$, $(1)(132) = (132)(1)$, $(123)(123) = (123)(123)$, $(123)(132) = (132)(123)$, $(132)(123) = (123)(132)$, $(132)(132) = (132)(132)$. Manifestly, $N(K)$ is a subgroup of G ; note, also, that $N(K)$ is Abelian.

$N(K)$ may consist of only one element of G or, if G is an Abelian group itself, all elements of G .

Th. 3.2.3.3 Given a subgroup G_1 of a group G , the complex K of all $x \in G_1$ such that $G_1^x = G_1$ is a subgroup of G . (Cf. Prob. 2.)

Df. 3.2.3.4 The complex K of Th. 3.2.3.3 is called the *centralizer* of G_1 in G , denoted by $C_{G_1}(K)$ or simply $C(K)$.

Example:

If, again, G is S_3 and $G_1 = \{(123), (132)\}$, then

$$C_{G_1}(K) = \{(123), (132)\} \quad \text{vs.} \quad N_{G_1}(K) = \{(1), (123), (132)\}$$

As is exemplified above and also explicit in Df. 3.2.3.2 and Df. 3.2.3.4, it is always the case that $C(K) \subseteq N(K)$ in any group.

Df. 3.2.3.5 If $G_1 = G$ in Th. 3.2.3.3, then $C(K)$ is called the *center* of G , and the elements of $C(K)$ are called the *central elements* of G .

Example:

If $G_1 = G$ in the examples of Df. 3.2.3.2 and 3.2.3.4 above, then $C(K) = N(K) = \{(1)\}$, as can be readily verified by the multiplication table of S_3 . (The center, of course, may be of more than one element; cf. Prob. 3 and also Th. 3.2.3.8.)

In general, $C(K)$ and $N(K)$ coincide if $G_1 = G$ — which is obvious from their definitions — just as they do when they consist of a single element. It is then self-explanatory to speak only of the normalizer or the centralizer of G if $G = G_1$.

Th. 3.2.3.6 The normalizer (or centralizer) of G_1 in G contains G_1 in itself. (Cf. Prob. 4.)

Th. 3.2.3.7 The number of the conjugates of a complex K contained in a group G is $(G : N(K))$, i.e. the index in G of the normalizer of K in G ; the index is also a divisor of the order of G . (Cf. Prob. 7.)

Th. 3.2.3.8 The center of a group G , where $G \neq \{e\}$, whose order is of p^n , where p is a prime and n any integer, is greater than the identity alone. (Cf. Prob. 11.)

Df. 3.2.3.9 If the order of every element except the identity of a group G is of a power of a prime p (as in Th. 3.2.3.8), G is called a *p-group*.

Th. 3.2.3.10 Any group G whose order is of p^2 , where p is a prime, is Abelian. (Cf. Prob. 12.)

Solved Problems

1. Prove Th. 3.2.3.1.

PROOF:

Since $x^{-1}gx = g$ for all $x \in G$ and all $g \in G_1$, it immediately follows that $g = xgx^{-1} = (x^{-1})^{-1}g(x^{-1})$. Hence, if $x \in K$, then also $x^{-1} \in K$. Moreover, if $y^{-1}gy = g$, then $(xy)^{-1}g(xy) = y^{-1}(x^{-1}gx)y = y^{-1}gy = g$. Hence $xy \in K$, satisfying the conditions of Th. 2.1.1.13, which proves K to be a subgroup of G .

2. Prove Th. 3.2.3.3.

PROOF:

The proof will be the same as above, only with a slight modification, since the only difference between these two theorems is a small (but important) stipulation: " $x \in G_1$ " vs. " $x \in G$ ".

The proof, however, may be given in a nominally different way, for instance, as follows:

If $x, y \in K$, i.e. $xg = gx$ and $yg = gy$ for all $x, y, g \in G_1$, then $(xy)g = x(yg) = x(gy) = (xg)y = (gx)y = g(xy)$, which implies $xy \in K$. Also, it follows from $xg = gx$ that $gx^{-1} = x^{-1}g$, which implies $x^{-1} \in K$. Hence, by Th. 3.1.1.13, K is proved to be a subgroup of G .

3. Find the center, the centralizers, and the normalizers of the octic group G of the square (cf. §3.1.2, Prob. 6): $0 = (a)(b)(c)(d)$, $1 = (abcd)$, $2 = (ac)(bd)$, $3 = (adcb)$, $4 = (ab)(cd)$, $5 = (ad)(bc)$, $6 = (ac)$, $7 = (bd)$.

Solution:

Let $K_1 = \{0\}$; then, using the multiplication table of G , K_1 is at once found to be the centralizer, in fact the center, of G in G itself. Obviously, then, K_1 is also the normalizer of G itself in G ; furthermore, since K_1 is a subgroup of G , K_1 is the normalizer of K_1 itself in G .

Likewise $K_2 = \{0, 2\}$, which is a subgroup of G , is the center, the centralizer, and the normalizer of G in G itself; also, K_2 is the normalizer of K_2 (as a subgroup) in G .

Let $K_3 = \{0, 1, 2, 3\}$, which, again, is a subgroup of G ; then the complex K_3 is the centralizer, and also the normalizer, of the subgroup K_3 in G ; moreover, K_3 may be considered the normalizer of K_1 or K_2 in G .

K_3 may be replaced by $K_4 = \{0, 2, 4, 5\}$ or $K_5 = \{0, 2, 6, 7\}$.

Likewise $K_6 = \{0, 4\}$ (or $K_7 = \{0, 5\}$) is the centralizer as well as the normalizer of K_6 (or K_7) in G or in K_4 ; K_6 (or K_7) may be considered the normalizer of K_1 in G or in K_4 or in K_6 itself.

$K_8 = \{0, 6\}$ (or $K_9 = \{0, 7\}$) has the same characteristics as K_6 (or K_7) with respect to centralizers and normalizers.

4. Prove Th. 3.2.3.6.

PROOF:

By definition, the normalizer (or the centralizer) consists of those elements of the whole group G (or a subgroup G_1 of G) that commute with every element of G_1 ; it must then contain G_1 in itself, since a subgroup always commutes with each of its elements.

This is true even if G_1 consists of a single element, since an element of a group always commutes with itself.

5. Given a cyclic group S generated by an element g of a group G , $N(S)$ contains all g^n , where $n \in I$, and the order of $N(S)$ is a multiple of the order of g .**PROOF:**

Since $(g^n)^{-1}g(g^n) = g$, i.e. $g(g^n) = (g^n)g$, it follows that g^n for any $n \in I$ must be contained in $N(S)$. Also, since the cyclic group S generated by g is now proved to be a subgroup of $N(S)$, the order of $N(S)$ must be divisible by the order of g .

Note. g belongs to the center of $N(S)$ of G ; hence $N(S)$ evidently contains the center of G .

6. There exists a 1-1 correspondence between the conjugates of an element g of a group G and the right cosets of the normalizer N of g .**PROOF:**

Let the right cosets of N be Nx , where $x \in G$, and also $ax \in xN$ for $a \in N$; then, since $a^{-1}ga = g$,

$$(ax)^{-1}g(ax) = x^{-1}a^{-1}gax = x^{-1}(a^{-1}ga)x = x^{-1}gx$$

Now, assume $x_1^{-1}gx_1 = x_2^{-1}gx_2$ when there are two distinct right-cosets x_1N and x_2N ; then

$$(x_2x_1^{-1})^{-1}gx_2x_1^{-1} = x_1x_2^{-1}gx_2x_1^{-1} = x_1x_1^{-1}gx_1x_1^{-1} = g$$

implying that $x_2x_1^{-1} \in N$, i.e. $x_2 \in Nx_1$, which in turn implies that $Nx_2 = Nx_1$, contradictory to the assumption. Hence the right-cosets are distinct, and to each right-coset of g there corresponds a distinct conjugate of g .

7. Prove Th. 3.2.3.7.

PROOF:

Since the index in G of the normalizer N of K is presumed, as usual, to be finite, G has a right decomposition

$$G = Nx_1 \cup Nx_2 \cup \dots \cup Nx_m$$

where there exists, as proved above (Prob. 6), a 1-1 correspondence: $Nx_i \leftrightarrow x_i^{-1}gx_i$. Hence the number of the conjugates of K must be $m = (G:N(K))$.

It also immediately follows, by Df. 3.2.2.9, that m is a divisor of the order of G .

8. Given the symmetric group $S_5 = \{a, b, c, d, e\}$, find the number of the conjugates of $p = (abcde)$ in S_5 .

Solution:

Since $(abcde) = (bcdea) = (cdeab) = (deabc) = (eabcd)$, the permutations through which p remain unvaried are

$$p^0 = p^5 = \begin{pmatrix} a & b & c & d & e \\ a & b & c & d & e \end{pmatrix}, \quad p = \begin{pmatrix} a & b & c & d & e \\ b & c & d & e & a \end{pmatrix}, \quad p^2 = \begin{pmatrix} a & b & c & d & e \\ c & d & e & a & b \end{pmatrix},$$

$$p^3 = \begin{pmatrix} a & b & c & d & e \\ d & e & a & b & c \end{pmatrix}, \quad p^4 = \begin{pmatrix} a & b & c & d & e \\ e & a & b & c & d \end{pmatrix}$$

Hence the order of the normalizer of p is 5 and, by Th. 3.2.3.7, the index is $5!/5 = 24$.

9. Find the number of the conjugates of $s = (abc)(def)$ in $S_6 = \{a, b, c, d, e, f\}$.

Solution:

Let $s_1 = (abc)$ and $s_2 = (def)$; then, since $(abc) = (bca) = (cab)$ and $(def) = (efd) = (fde)$, the set P of all permutations which do not change s_1 and s_2 are:

$$s_1^0 s_2^0, s_1^1 s_2^0, s_1^2 s_2^0; s_1^0 s_2^1, s_1^1 s_2^1, s_1^2 s_2^1; s_1^0 s_2^2, s_1^1 s_2^2, s_1^2 s_2^2$$

Since the normalizer N of s in S_6 , of which N is obviously a subset, also contains another subset Q in which s_1 and s_2 interchange themselves, it follows that $N = P \cup Qq$, where $q = (14)(25)(36)$. But the order of P is 9 and, consequently, that of Qq is 9. Hence the order of N is 18, and the number of the conjugates of s in S_6 is $6!/18 = 40$.

10. Generalize the result of Prob. 8-9.

PROOF:

Let s in S_n consist of c_1 cycles of length 1, c_2 cycles of length 2, ..., c_k cycles of length k ; then, obviously,

$$n = 1c_1 + 2c_2 + \dots + kc_k$$

Now, the permutations of n distinct letters taken all at a time is $n!$, and since a cycle of length c_i may be freely permuted without changing s as a whole, the same permutation occurs $c_1! c_2! \dots c_k!$ times if all $n!$ cases are considered. Moreover, since a cycle of length c_i may be written in c_i different ways and allows thus i^{c_i} equivalent cases, each permutation has been repeatedly considered $1^{c_1} 2^{c_2} \dots k^{c_k}$ times. Hence the number of distinct permutations is

$$n! / (1^{c_1} c_1! 2^{c_2} c_2! \dots k^{c_k} c_k!)$$

which is also the number of the conjugates of s in S_n .

11. The center C of a group G is a subgroup of G and, if G is Abelian, is G itself.

PROOF:

For every $g \in G$, $x, y \in C$ implies $g^{-1}xg = x$ and $g^{-1}yg = y$, which in turn implies $g^{-1}xgg^{-1}x^{-1}g = g^{-1}xx^{-1}g = e$. Hence $g^{-1}x^{-1}g = (g^{-1}xg)^{-1} = x^{-1}$, and consequently $g^{-1}(x^{-1}y)g = g^{-1}x^{-1}gg^{-1}yg = x^{-1}y \in C$, proving that C is a subgroup of G .

If G is Abelian, then every element of G commutes with all other elements of G ; hence C coincides with G .

(E.g. S_2 , which consists of e and a transposition, constituting an Abelian group, is the center of S_2 itself. If $n \geq 3$, however, the center of S_n consists of $\{e\}$ alone, just as any non-commutative simple groups (cf. Df. 3.2.5.2a).

12. Prove Th. 3.2.3.8.

PROOF:

Decompose G into a class equation

$$G = C_1 \cup C_2 \cup \dots \cup C_m$$

and let a_i be the number of elements in C_i ; then

$$p^n = a_1 + a_2 + \dots + a_m$$

Assume that some of C_i are the identity (which is always a class in itself and naturally belongs to the center of G). Then $k \geq 1$, k denoting the number of such identity classes. This implies, by Th. 3.2.3.7, that p^n must be divisible by $(m - k)$, the number of non-identity classes, i.e.,

$$p^n = k + (m - k) = k + ip, \quad i = 1, 2, \dots, n$$

since p^n is obviously divisible by p itself. Hence k also must be a multiple of p , i.e. $k > 1$, completing the proof.

13. Prove Th. 3.2.3.10.

PROOF:

The order q of the center C of a group G is greater than 1, as has been proved by Th. 3.2.3.8. Also, since C is a subgroup of G , the order p^2 of G must be divisible by q . Hence $q = p^2$ or $q = p$.

If $q = p^2$, then $C = G$ and, by definition, every element which belongs to G becomes permutable with every element of G . Hence G must be an Abelian group.

If $q = p$, assume $a \in G$ and $a \notin C$. But, then, since the centralizer C' of a must contain C and also all powers of a , the order q' of C' is greater than p , which implies $q' = p^2$ (for p^2 must be divisible by q'). Hence a must belong to C , contradictory to the assumption.

Hence it must be the case that $q = p^2$, which proves G to be Abelian.

*§3.2.4 Endomorphism and Automorphism

Df. 3.2.4.1 A homomorphism (cf. Df. 3.1.3.1-2) of a group G onto G itself or into a subgroup G_1 of G is called an *endomorphism* (cf. Df. 2.2.2.10) of G or an *operator* on G .

Stated otherwise, a transformation $T: G \rightarrow G$ (or G_1) is an endomorphism of G if $T(ab) = T(a)T(b)$ for every $a, b \in G$; i.e. there exists a correspondence: $a \rightarrow a'$, $b \rightarrow b'$, ..., for $a, b, \dots \in G$ and also a', b', \dots (not necessarily distinct) $\in G$, such that $a \rightarrow a'$ and $b \rightarrow b'$ imply $ab \rightarrow a'b'$.

An endomorphism of G is logically equivalent to an *operator* on G which in effect is a mapping $f: f(a) = a', f(b) = b', \dots$ which entails the distributive property: $f(ab) = f(a)f(b) = a'b'$.

Df. 3.2.4.2 Every group G has at least two operators: the *identity operator* I such that $I(a) = a$ for any $a \in G$, and the *null operator* $N(a) = e$ which is the identity element of G .

Th. 3.2.4.3 Every endomorphism maps every subgroup G_i of G onto or into a subgroup $f(G_i)$ of G . (Cf. Prob. 1.)

Th. 3.2.4.4 The product of two endomorphisms is in itself an endomorphism. (Cf. Prob. 2.)
Operators as such, when put together, are associative. (Cf. Prob. 3.)

Df. 3.2.4.5 If an endomorphism of G is 1-1, then it is an *automorphism* of G .

Stated otherwise, an automorphism of G is an isomorphism of G into and onto G itself, viz. a transformation $T: G \leftrightarrow G$ where $T(ab) = T(a)T(b)$ for every $a, b \in G$; i.e. there exists a 1-1 correspondence between G and G itself such that, for every $a, a', b, b', \dots \in G$, $a \leftrightarrow a'$ and $b \leftrightarrow b'$ imply $ab \leftrightarrow a'b'$.

In general, if $T(a) = T(b)$ implies $a = b$ for every $a, b \in G$, the endomorphism is an isomorphism which in turn, in a finite group, is necessarily an automorphism. (This is not the case in an infinite group F where F may be isomorphic to a proper subgroup; e.g. $x \rightarrow nx$ for some integer n is an endomorphism which in itself is an isomorphism of the additive group of I , the set of all integers, but this endomorphism cannot be an automorphism. Conversely, however, every isomorphism between a group and one of its proper subgroup is also an endomorphism, but not an automorphism.)

Th. 3.2.4.6 The composite (or product) of two automorphisms is also an automorphism. (Cf. Prob. 4.)

Th. 3.2.4.7 $A(G)$, which denotes the class of all automorphisms of a group G , forms a group. (Cf. Prob. 5.)

Note. $A(G)$ is in fact a subgroup of $T(G)$, the class of all possible transformations of G , one of which is exemplified below.

Th. 3.2.4.8 The transformation C of the conjugates $g \leftrightarrow x^{-1}gx$ for some x and any element g of a group G is an automorphism of G . (Cf. Prob. 6.)

Df. 3.2.4.9 The automorphism C of the form $g \leftrightarrow x^{-1}gx$ in Th. 3.2.4.8 is called an *inner automorphism*.

Note. All inner automorphisms of an Abelian group are necessarily reduced to the identical transformation. (Cf. Prob. 7.)

Th. 3.2.4.10 The inner automorphisms of a group G form a subgroup of $A(G)$, denoted by $C(G)$. (Cf. Th. 3.2.3.8 and Prob. 8.)

Df. 3.2.4.11 Any automorphism which does not belong to $C(G)$ is an *outer automorphism*.

In general, an outer automorphism is an automorphism which is not mapped by a single element; e.g. the mapping in the four group V_4 : $e \leftrightarrow e (= a^2 = b^2)$, $a \leftrightarrow b$, $b \leftrightarrow a$, $ab \leftrightarrow ba$, is an outer automorphism.

Solved Problems

1. Prove Th. 3.2.4.3.

PROOF:

G1. Let $a_i, b_i \in G_i$; then, by Df. 3.2.3.1, $a_i \rightarrow a'_i = f(a_i)$ and $b_i \rightarrow b'_i = f(b_i)$, where $f(a_i), f(b_i) \in f(G_i)$, imply $a_i b_i \in G_i \rightarrow f(a_i b_i) = f(a_i) f(b_i) = a'_i b'_i \in f(G_i)$.

G2. If $a_i, b_i, c_i \in G$, then $a_i(b_i c_i) = (a_i b_i) c_i \rightarrow f(a_i(b_i c_i)) = f((a_i b_i) c_i) \rightarrow f(a_i) f(b_i) f(c_i) = (f(a_i) f(b_i)) f(c_i) \rightarrow a'_i(b'_i c'_i) = (a'_i b'_i) c'_i$, where $a'_i, b'_i, c'_i \in f(G_i)$.

G3: Since $f(e) = e$ and $f(a_i) = f(a_i e) = f(a_i) f(e)$, it follows that $a'_i e = a'_i$.

G4: Since $e = f(e) = f(a_i a_i^{-1}) = f(a_i) f(a_i^{-1})$, it follows that $a'_i a_i'^{-1} = e$.

2. Prove Th. 3.2.4.4.

PROOF:

Given two endomorphisms E_1 and E_2 of G , defined as $E_1: a' \in G \rightarrow E_1(a') = a'' \in G$, $E_2: a \in G \rightarrow E_2(a) = a' \in G$, respectively, it follows that $E_1E_2: a \in G \rightarrow E_1E_2(a) = E_1(E_2(a)) = E_1(a') = a'' \in G$.

Likewise

$$b \in G \rightarrow E_1E_2(b) = E_1(E_2(b)) = E_1(b') = b'' \in G$$

$$\text{and } ab \in G \rightarrow E_1E_2(ab) = E_1(E_2(ab)) = E_1(E_2(a)E_2(b)) = E_1(a'b') = E_1(a')E_1(b') = a''b'' \in G$$

Hence the composite (or product) of E_1 and E_2 is also an endomorphism of G .

3. The composition (or multiplication) of endomorphisms is associative.

PROOF:

Let, as in Prob. 2 above, $E_3: a \in G \rightarrow E_3(a) = a' \in G$, $E_2: a' \in G \rightarrow E_2(a') = a'' \in G$, $E_1: a \in G \rightarrow E_1(a') = a''' \in G$. Then $E_1(E_2E_3) = E_1(E_2E_3)(a) = E_1(E_2(E_3(a))) = E_1(E_2(a')) = E_1(a'') = a'''$ and likewise $(E_1E_2)E_3 = (E_1E_2)E_3(a) = (E_1E_2)(a') = E_1(E_2(a')) = E_1(a'') = a'''$. Hence $E_1(E_2E_3) = (E_1E_2)E_3$.

4. Prove Th. 3.2.4.6.

PROOF:

Given two automorphisms A_1 and A_2 of G , defined as $A_2: a \in G \leftrightarrow A_2(a) = a' \in G$ and $A_1: a' \in G \leftrightarrow A_1(a') = a'' \in G$; then, replacing only “ \rightarrow ” of Prob. 2 above by “ \leftrightarrow ”, the desired result is obtained, viz. $a \in G \leftrightarrow A_1A_2(a) = a'' \in G$ and $b \in G \leftrightarrow A_1A_2(b) = b'' \in G$ imply $ab \in G \leftrightarrow A_1A_2(ab) = a''b'' \in G$.

5. Prove Th. 3.2.4.7.

PROOF:

G1, the closure property, is provided by Th. 3.2.3.6.

G2 for $\mathbf{A}(G)$ is obtained from Prob. 3 above with a slight modification, viz. replacing “ \rightarrow ” by “ \leftrightarrow ”.

G3. $I \subset \mathbf{A}(G)$, since the identity transformation $I(A_i) = A_i$, $A_i \subset \mathbf{A}(G)$, is clearly an automorphism.

G4. By Df. 3.2.4.5, $A_i \leftrightarrow \mathbf{A}^{-1}(A_i) = A_i^{-1}$ and $A_j \leftrightarrow \mathbf{A}^{-1}(A_j) = A_j^{-1}$ imply

$$\mathbf{A}^{-1}(A_iA_j) = \mathbf{A}^{-1}(\mathbf{A}\mathbf{A}^{-1}(A_iA_j)) = \mathbf{A}^{-1}(\mathbf{A}(\mathbf{A}^{-1}(A_i)\mathbf{A}^{-1}(A_j))) = \mathbf{A}^{-1}(A_i)\mathbf{A}^{-1}(A_j) = A_i^{-1}A_j^{-1}$$

Hence $\mathbf{A}(G)$ forms a group.

6. Prove Th. 3.2.4.8.

PROOF:

For any $g_1, g_2 \in G$, $g_1 \leftrightarrow C(g_1) = x^{-1}g_1x$ and $g_2 \leftrightarrow C(g_2) = x^{-1}g_2x$ do imply $g_1g_2 \leftrightarrow C(g_1g_2) = (x^{-1}g_1x)(x^{-1}g_2x) = x^{-1}g_1g_2x$. Hence the mapping C is an automorphism.

7. In an Abelian group all inner automorphisms become the identity transformation.

PROOF:

Let $A_i: g_i \leftrightarrow A_i(g_i) = x^{-1}g_ix$ for some x and every $g_i \in G$; then, if G is Abelian,

$$A_i(g_i) = x^{-1}g_ix = x^{-1}xg_i = g_i = I(g_i)$$

8. Prove Th. 3.2.4.10.

PROOF:

Let $g_1, g_2, g_3 \in G$ and $C_1: g_1 \leftrightarrow C_1(g_1) = x^{-1}g_1x$, $C_2: g_2 \leftrightarrow C_2(g_2) = x^{-1}g_2x$, $C_3: g_3 \leftrightarrow C_3(g_3) = x^{-1}g_3x$; then

G1. $g_1g_2 \leftrightarrow C_1(g_1)C_2(g_2) = x^{-1}(g_1g_2)x$ (cf. Prob. 6).

G2. By **G1** immediately above, $C_1(C_2C_3) = x^{-1}(g_1(g_2g_3))x = x^{-1}((g_1g_2)g_3)x = (C_1C_2)C_3$.

G3. $g_1e = g_1 \leftrightarrow C_1(g_1)C_1(e) = (x^{-1}g_1x)(x^{-1}ex) = x^{-1}(g_1e)x = x^{-1}g_1x = C_1(g_1)$.

G4. $g_1g_1^{-1} = e \leftrightarrow C_1(g_1)C_1(g_1^{-1}) = (x^{-1}g_1x)(x^{-1}g_1^{-1}x) = x^{-1}(g_1g_1^{-1})x = x^{-1}ex = e$.

Hence the class of inner automorphisms forms a group, denoted by $\mathbf{C}(G)$; moreover, since **G1** above of $\mathbf{C}(G)$ satisfies Th. 3.2.1.1, $\mathbf{C}(G)$ is a subgroup of $\mathbf{A}(G)$.

Second proof. Since, by Th. 3.2.4.7, $\mathbf{A}(G)$ is already a group and $\mathbf{C}(G)$ is obviously a non-empty subclass of $\mathbf{A}(G)$, by Df. 3.2.4.9, $\mathbf{C}(G)$ is a subgroup of $\mathbf{A}(G)$ iff $C_i(G), C_j(G) \subset \mathbf{C}(G)$ implies $C_i^{-1}(G)C_j(G) \subset \mathbf{C}(G)$. This is exactly the case, since for some x and any $g \in G$,

$$C_i^{-1}(G)C_j(G) = (x^{-1}g_ix)^{-1}(x^{-1}g_jx) = (x^{-1}g_ix)(x^{-1}g_j^{-1}x) = x^{-1}g_ig_j^{-1}x = x^{-1}g_kx = C_k(G) \subset \mathbf{C}(G)$$

The converse is obvious, completing the proof.

9. Find, if any, inner and outer automorphisms in (i) the cyclic group of order 3, and (ii) the symmetric group of order 3.

Solution:

- (i) $x^3 = e$ is obviously an inner automorphism, since any identity transformation is an inner automorphism, as can be readily verified, but no other elements have the same kind of transformations. Between x and x^2 , however, there exists an automorphism S of squaring, $x \leftrightarrow S(x) = x^2$ and $x^2 \leftrightarrow S(x^2) = x$, which together imply $xx^2 = e \leftrightarrow S(xx^2) = S(x)S(x^2) = x^2x = e$. Since S is evidently not an inner automorphism, it must be an outer automorphism, satisfying the definition (cf. Df. 3.2.3.11).
- (ii) S_3 consists of six inner automorphisms, thus denying any occurrence of outer automorphisms; viz., for $x^{-1}sx \leftrightarrow s$,

(i) $x = (1)$:	(ii) $x = (12)$:	(iii) $x = (13)$:
$(1) \leftrightarrow (1)$	$(1) \leftrightarrow (1)$	$(1) \leftrightarrow (1)$
$(12) \leftrightarrow (12)$	$(12) \leftrightarrow (12)$	$(12) \leftrightarrow (12)$
$(13) \leftrightarrow (13)$	$(13) \leftrightarrow (13)$	$(13) \leftrightarrow (13)$
$(23) \leftrightarrow (23)$	$(23) \leftrightarrow (23)$	$(23) \leftrightarrow (23)$
$(123) \leftrightarrow (123)$	$(123) \leftrightarrow (123)$	$(123) \leftrightarrow (123)$
$(132) \leftrightarrow (132)$	$(132) \leftrightarrow (132)$	$(132) \leftrightarrow (132)$

Likewise for $x = (23), (123), (132)$.

10. If a group G is defined by the following multiplicative rules,

$$a^2 = b^2 = c^2 = e, \quad ab = ba = c, \quad bc = cb = a, \quad ca = ac = b$$

G has then only one inner automorphism and five outer automorphisms.

PROOF:

G is octic and also Abelian, as is explicitly defined above; hence, by Prob. 7, G has only one inner automorphism which is the identity transformation.

Other automorphisms of G , which by definition are outer automorphisms, are as follows:

$$\begin{aligned} A_1: & A_1(a) = b, \quad A_1(b) = a, \quad A_1(c) = c \\ A_2: & A_2(a) = c, \quad A_2(b) = b, \quad A_2(c) = a \\ A_3: & A_3(a) = a, \quad A_3(b) = c, \quad A_3(c) = b \\ A_4: & A_4(a) = b, \quad A_4(b) = c, \quad A_4(c) = a \\ A_5: & A_5(a) = c, \quad A_5(b) = a, \quad A_5(c) = b \end{aligned}$$

* * * * *

G of Prob. 10 is an octic group, but not the octic group of the square; other four octic groups, vs. the Dihedral group, D_4 , are shown below.

1. Dihedral Group, characterized by $a^4 = b^2 = e$, $ba = a^3b$ (which, incidentally, has four inner and four outer automorphisms):

	e	a	a^2	a^3	b	ab	a^2b	a^3b
e	e	a	a^2	a^3	b	ab	a^2b	a^3b
a	a	a^2	a^3	e	ab	a^2b	a^3b	b
a^2	a^2	a^3	e	a	a^2b	a^3b	b	ab
a^3	a^3	e	a	a^2	a^3b	b	ab	a^2b
b	b	a^3b	a^2b	ab	e	a^3	a^2	a
ab	ab	b	a^3b	a^2b	a	e	a^3	a^2
a^2b	a^2b	ab	b	a^3b	a^2	a	e	a^3
a^3b	a^3b	a^2b	ab	b	a^3	a^2	a	e

2. Quaternion Group: $a^4 = e$, $a^2 = b^2$, $ba = a^3b$

	e	a	a^2	a^3	b	ab	a^2b	a^3b
e	e	a	a^2	a^3	b	ab	a^2b	a^3b
a	a	a^2	a^3	e	ab	a^2b	a^3b	b
a^2	a^2	a^3	e	a	a^2b	a^3b	b	ab
a^3	a^3	e	a	a^2	a^3b	b	ab	a^2b
b	b	a^3b	a^2b	ab	a^2	a	e	a^3
ab	ab	b	a^3b	a^2b	a^3	a^2	a	e
a^2b	a^2b	ab	b	a^3b	e	a^3	a^2	a
a^3b	a^3b	a^2b	ab	b	a	e	a^3	a^2

3. $a^8 = e$

	e	a	a^2	a^3	a^4	a^5	a^6	a^7
e	e	a	a^2	a^3	a^4	a^5	a^6	a^7
a	a	a^2	a^3	a^4	a^5	a^6	a^7	e
a^2	a^2	a^3	a^4	a^5	a^6	a^7	e	a
a^3	a^3	a^4	a^5	a^6	a^7	e	a	a^2
a^4	a^4	a^5	a^6	a^7	e	a	a^2	a^3
a^5	a^5	a^6	a^7	e	a	a^2	a^3	a^4
a^6	a^6	a^7	e	a	a^2	a^3	a^4	a^5
a^7	a^7	e	a	a^2	a^3	a^4	a^5	a^6

4. $a^4 = b^2 = e$

	e	a	a^2	a^3	b	ab	a^2b	a^3b
e	e	a	a^2	a^3	b	ab	a^2b	a^3b
a	a	a^2	a^3	e	ab	a^2b	a^3b	b
a^2	a^2	a^3	e	a	a^2b	a^3b	b	ab
a^3	a^3	e	a	a^2	a^3b	b	ab	a^2b
b	b	ab	a^2b	a^3b	e	a	a^2	a^3
ab	ab	a^2b	a^3b	b	a	a^2	a^3	e
a^2b	a^2b	a^3b	b	ab	a^2	a^3	e	a
a^3b	a^3b	b	ab	a^2b	a^3	e	a	a^2

5. $a^2 = b^2 = c^2$

	e	a	b	c	ab	ac	bc	abc
e	e	a	b	c	ab	ac	bc	abc
a	a	e	ab	ac	b	c	abc	bc
b	b	ab	e	bc	a	abc	c	ac
c	c	ac	bc	e	abc	a	b	ab
ab	ab	b	a	abc	e	bc	ac	c
ac	ac	c	abc	a	bc	e	ab	b
bc	bc	abc	c	b	ac	ab	e	a
abc	abc	bc	ac	ab	c	b	a	e

*§3.2.5 Normal Subgroups

Df. 3.2.5.1 The *conjugate subgroups* (cf. Df. 3.2.2.14, Df. 3.2.4.9-10) of a group G are of the form $g^{-1}G_i g$, for every $g \in G$, where G_i is any subgroup of G .

Example:

A subgroup G_1 conjugate to a subgroup $G_2 = \{(1), (123), (123)^2\}$ under the symmetric group S_4 may be obtained by a transforming element, say, (14) : $(14)^{-1}(1)(14) = (1)$; $(14)^{-1}(123)(14) = (423)$; $(14)^{-1}(132)(14) = (432)$. (Cf. Prob. 4.)

Df. 3.2.5.2 Any conjugate subgroup G_i of a group G is called *normal* (or *self-conjugate* or *invariant*) iff $G_i^x = G_i$, i.e. $x^{-1}G_i x = G_i$, for all $x \in G$.

Stated otherwise: a subgroup G_i of a group G is normal (in G) iff G_i remains the same under all inner automorphisms of G .

Example:

The subgroup $\{(1), (13)(24)\}$ of the dihedral group (cf. Prob. 5 below) is normal; so are G itself and $\{e\}$, as is self-explanatory in Df. 3.2.5.2 itself.

Hence the following ramification:

Df. 3.2.5.2a A group containing none of proper normal subgroups, except itself and $\{e\}$, is called a *simple group*. (Cf. Prob. 24.)

All Abelian groups are necessarily normal, since $x^{-1}gx = gx^{-1}x = g$ in any Abelian group; so is every subgroup of an Abelian group. But not conversely, since there does exist a family of non-Abelian groups, every subgroup of which is normal. Hence, again, a ramification as follows:

Df. 3.2.5.2b A non-Abelian group, all of whose subgroups are normal, is called a *Hamiltonian group*.

One member of this family, viz. the *quaternion group*, has already been observed (cf. §3.2.4, Prob. 10 note, and also Prob. 23 below), but other members at large are beyond the scope of the present study.

In general, the normalizer $N(G_1)$ of a subgroup G_1 of a group G is plainly the maximal subgroup of G in which G_1 is normal; thus, the normalizer of G_1 is G itself iff G_1 is normal in G , i.e. $N(G_1) = G$ (cf. §3.2.3, Prob. 3, and also Prob. 5-6 below). As is obvious from their definitions, the centralizer $C(G_1)$ is a normal subgroup of $N(G_1)$.

Normal subgroups may be characterized otherwise, e.g. as in the following theorems.

Th. 3.2.5.3 A subgroup G_1 of a group G is normal iff $gG_1 = G_1g$ for every $g \in G$. (Cf. Prob. 8.)

Th. 3.2.5.4 Any subgroup of index 2 (cf. Df. 3.2.2.9) is necessarily normal. (Cf. Prob. 10.)

Stated otherwise: every subgroup G_i of G is normal if G_i has only one other coset; all g_j such that $g_j \in G$ and $g_j \notin G_i$ will then form the right and left coset of G_i in G . This is most clearly exemplified in the following theorem.

Th. 3.2.5.5 The alternating group A_n of the symmetric group S_n is normal. (Cf. Df. 3.1.2.18, and Prob. 11 below.)

Example:

The symmetric group S_4 has the alternating group

$$A_4 = \{(1), (123), (123)^2, (124), (124)^2, (134), (134)^2, (234), (234)^2, (12)(34), (13)(24), (14)(23)\}$$

which is normal, as can be individually verified without difficulty. Note, also, that A_4 itself has a normal subgroup, viz. the Vierergruppe (cf. §3.1.2, Prob. 2, 12; §3.1.3, Prob. 12),

$$V_4 = \{(1), (12)(34), (13)(24), (14)(23)\}$$

Th. 3.2.5.6 For every complex K of a group G and every normal subgroup N of G , $KN = NK$. (Cf. Prob. 14.)

Th. 3.2.5.7 For every subgroup S of a group G and every normal subgroup N of G , $SN = NS$ is a normal subgroup of G . (Cf. Prob. 15.)

Th. 3.2.5.8 (First Theorem on Homomorphism). If a group G' is a homomorph of a group G under a homomorphism H , then a complex K of every element k , $k \in G$, whose homomorph is the identity e' , where $e' \in G'$, forms a normal subgroup of G . (Cf. Prob. 16.)

Df. 3.2.5.9 The normal subgroup of Th. 3.2.5.8 is called the *kernel* of the homomorphism H .

Example:

The homomorphism H of the alternating group A_4 onto the cyclic group of order 3 (cf. §3.2.1, Prob. 18) has the kernel: $\{(1), (12)(34), (13)(24), (14)(23)\}$.

Th. 3.2.5.10 Any two elements of a group G have the same image in a homomorph G' of G iff they are in the same coset of the kernel S . (Cf. Prob. 17.)

Th. 3.2.5.11 Th. 3.2.5.10 implies $(xS)(yS) = xyS$, where $x, y \in G$. (Cf. Prob. 18.)

This theorem holds, more generally, also for any normal subgroup N of a group G , as is quite explicit in the proof.

Th. 3.2.5.12 If there exists a homomorphism H of a group G onto or into a group G' , and if G_1 is a subgroup of G , then $H(G_1)$ is a subgroup of G' , i.e. $H(G)$; if G_1 is a normal subgroup of G , so is then $H(G_1)$ of $H(G)$. (Cf. Prob. 19.)

Solved Problems

1. If a subgroup G_1 of G is Abelian, so is a subgroup G_2 conjugate to G_1 under G .

PROOF:

Let $a, b \in G_1$; then $ab = ba$ and also $(g^{-1}ag), (g^{-1}bg) \in G_2$ for every $g \in G$. Now, $(g^{-1}ag)(g^{-1}bg) = g^{-1}abg = g^{-1}bag = g^{-1}bgg^{-1}ag = (g^{-1}bg)(g^{-1}ag)$. Hence G_2 is also Abelian.

2. The order and the index in G of a subgroup G_1 equal those of a subgroup G_2 conjugate to G_1 under G .

PROOF:

Let the order of G_1 be m , i.e. $a_1, a_2, \dots, a_m \in G_1$; then, for every $g \in G$,

$$G_2 = g^{-1}G_1g = \{g^{-1}a_1g, g^{-1}a_2g, \dots, g^{-1}a_mg\}$$

each element of which must be proved to be distinct.

If $g^{-1}a_i g = g^{-1}a_j g$, $1 \leq i \neq j \leq m$, then

$$a_i = gg^{-1}a_i gg^{-1} = gg^{-1}a_j gg^{-1} = a_j$$

contradictory to the assumption. Hence the elements of G_2 are all distinct, providing a 1-1 correspondence: $a_i \leftrightarrow g^{-1}a_i g$.

Thus if the order of G is n , the index in G of G_1 is n/m , which is also the index in G of G_2 .

3. A subgroup G_2 conjugate to a cyclic subgroup G_1 of a group G is also cyclic; the generator of G_2 is $g^{-1}ag$ for every $g \in G$ if the generator of G_1 is a .

PROOF:

The problem is to establish a 1-1 correspondence: $a^n \leftrightarrow (g^{-1}ag)^n = g^{-1}a^ng$.

If $n = 1$, then clearly $a \leftrightarrow g^{-1}ag$.

Suppose $a^k \leftrightarrow (g^{-1}ag)^k = g^{-1}a^kg$ for $n = k$; then, for $n = k + 1$,

$$g^{-1}a^{(k+1)}g = g^{-1}a^k(gg^{-1})ag = (g^{-1}a^kg)g^{-1}ag = (g^{-1}ag)^{(k+1)}$$

Also, if $n = 0$, then $g^{-1}a^0g = g^{-1}eg = e = (g^{-1}ag)^0$, and if $n = -1$, then $g^{-1}a^{-1}g = (g^{-1}ag)^{-1}$ since $(g^{-1}ag)(g^{-1}a^{-1}g) = e$; for $n < 0$ in general, since $-n > 0$,

$$g^{-1}a^ng = g^{-1}(a^{-n})^{-1}g = (g^{-1}a^{-n}g)^{-1} = ((g^{-1}ag)^{-n})^{-1} = (g^{-1}ag)^n$$

Hence it is always the case that $(g^{-1}ag)^n = g^{-1}a^ng$ for any integer n and, since Prob. 2 above assures the 1-1 correspondence $a^n \in G_1 \leftrightarrow g^{-1}a^ng \in G_2$, there also exists the 1-1 correspondence

$$a^n \in G_1 \leftrightarrow (g^{-1}ag)^n \in G_2$$

completing the proof.

4. Find all subgroups conjugate to a cyclic group C generated by (123) under the symmetric group $S_4 = \{1, 2, 3, 4\}$.

Solution:

The cyclic group C consists of $(1), (123), (123)^2 = (132)$, and $d^{-1}Cd = C$ for every transforming element $d \in D = \{(14), (24), (34), (4)\}$. Hence the conjugate subgroups are: $(14)^{-1}C(14) = \{(1), (423), (432)\}$, $(24)^{-1}C(24) = \{(1), (143), (134)\}$, $(34)^{-1}C(34) = \{(1), (124), (142)\}$, $(4)^{-1}C(4) = \{(1), (123), (132)\}$.

5. Find the normal subgroups of the octic group of the square.

Solution:

Referring to §3.2.3, Prob. 3, the subgroup $\{0\}$ is at once found to be normal; $\{0, 2\}$ is also normal.

Note. The normalizer N , e.g. $\{0, 1, 2, 3\}$, is evidently the largest of all subgroups of G which contain $\{0, 1, 2, 3\}$ as a normal subgroup. Likewise, $N(G_1)$ of $G_1 = \{0, 2\}$ is the maximal subgroup of G in which G_1 is normal.

6. Let A be a subgroup of a group B which in turn is a subgroup of G ; then, if A is normal in G , so is A in B .

PROOF:

By hypothesis, $g^{-1}ag = a$ for every $a \in A$ and every $g \in G$. Since B is a subgroup of G , every $b \in B$ implies every $b \in G$; hence $b^{-1}ab = a$, i.e. A is normal in B .

7. If two subsets G_1 and G_2 of a group G are normal, so are G_1G_2 and $G_1 \cap G_2$.

PROOF:

By hypothesis, $g^{-1}G_1g = G_1$ (or $G_1 = gG_1g^{-1}$) and $g^{-1}G_2g = G_2$ (or $G_2 = gG_2g^{-1}$).

(i) $g^{-1}G_1G_2g = g^{-1}(gG_1g^{-1})(gG_2g^{-1})g = G_1G_2$, i.e. G_1G_2 is normal in G .

(ii) Obviously, $g^{-1}(G_1 \cap G_2)g \subseteq (g^{-1}G_1g) \cap (g^{-1}G_2g)$. Conversely, if $a \in (g^{-1}G_1g) \cap (g^{-1}G_2g)$, then $a = g^{-1}bg$ implies $b \in (G_1 \cap G_2)$ and, consequently, $a \in g^{-1}(G_1 \cap G_2)g$. Hence $(g^{-1}G_1g) \cap (g^{-1}G_2g) \subseteq g^{-1}(G_1 \cap G_2)g$.

Hence, altogether, $g^{-1}(G_1 \cap G_2)g = (g^{-1}G_1g) \cap (g^{-1}G_2g) = G_1 \cap G_2$, i.e. the meet of G_1 and G_2 is also normal in G .

8. Prove Th. 3.2.5.3.

PROOF:

If $gG_1 = G_1g$ for every $g \in G$, then $g^{-1}G_1g = G_1$, which complies with Df. 3.2.5.2.

Conversely, if G_1 is normal in G , i.e. $g^{-1}G_1g = G_1$, then $gG_1 = g(g^{-1}G_1g) = G_1g$.

9. A subgroup G_1 of a group G is normal in G if $g^{-1}G_1g \subseteq G_1$ for every $g \in G$.

PROOF:

Since $g^{-1}G_1g = G_1$ for every $g \in G$, replace g by $g^{-1} \in G$, and

$$g^{-1}G_1g \rightarrow (g^{-1})^{-1}G_1(g^{-1}) = gG_1g^{-1} \subseteq G_1$$

Hence

$$G_1 = g^{-1}(gG_1g^{-1})g \subseteq gG_1g^{-1}$$

and together with the original hypothesis $g^{-1}G_1g \subseteq G_1$, it now follows that $G_1 = gG_1g^{-1}$, i.e. $gG_1 = G_1g$, completing the proof.

Note. The stipulation " $g^{-1}G_1g \subseteq G_1$ ", when used for defining G_1 to be a normal subgroup, is manifestly weaker than " $g^{-1}G_1g = G_1$ "; but, as proved above, the latter is deducible from the former, and the simpler form of the former may often be used for finding normal subgroups (cf. Prob. 16, 19 below).

10. Prove Th. 3.2.5.4.

PROOF:

The left-decomposition of G is, by hypothesis,

$$G = G_1 \cup pG_1$$

where p is an odd permutation in S_3 and pG_1 has no element in common with G_1 , while the right-decomposition is

$$G = G_1 \cup G_1p$$

Hence $pG_1 = G_1p$ and, by Th. 3.2.5.3, G_1 must be normal in G .

11. Prove Th. 3.2.5.5.

PROOF:

By Df. 3.1.2.18 and Th. 3.1.2.19, A_n is a subgroup of S_n which consists of A_n itself and but one coset of A_n ; hence, by Th. 3.2.5.4, A_n is necessarily normal in S_n .

Second proof (without resorting to Th. 3.2.5.4). Let p be an even permutation, i.e. $p = p_1p_2 \dots p_n$ where n is an even number and p_i is a transposition; then, for any permutation $q \in S_n$,

$$q^{-1}pq = q^{-1}(p_1p_2 \dots p_n)q = (q^{-1}p_1q)(q^{-1}p_2q) \dots (q^{-1}p_nq)$$

where $q^{-1}p_iq$ is evidently a transposition. Hence $q^{-1}pq$ is again an even permutation and, A_n being the subgroup of all even permutations in S_n , $q^{-1}A_nq \subseteq A_n$.

Conversely, for any even permutation p' , $qp'q^{-1} = (q^{-1})^{-1}p'q^{-1}$ is an even permutation; hence $p' = q^{-1}(qp'q^{-1})q \in q^{-1}A_nq$, which implies $A_n \subseteq q^{-1}A_nq$.

Hence $A_n = q^{-1}A_nq$, which proves A_n to be normal.

12. Any proper subgroup of S_3 which is not an alternating group is not a normal subgroup.

PROOF:

Of all subgroups of S_3 , $A_3 = \{(1), (123), (132)\}$ is an alternating group which, by Th. 3.2.5.5, is also normal.

Other proper subgroups: $\{(1), (12)\}$, $\{(1), (13)\}$, $\{(1), (23)\}$, however, are conjugate to each other; hence they cannot be normal, a normal subgroup being a subgroup conjugate only to itself.

13. A normal subgroup N of a group G is a class of conjugates, i.e. a join of one or more conjugate sets in G .

PROOF:

If $n_1 \in N$, then $g^{-1}n_1g \subseteq g^{-1}Ng = N$ for every $g \in G$; i.e. any element which belongs to the complex $g^{-1}n_1g$ also belongs to N . This holds for each $n_i \in N$, $i = 1, 2, \dots, k$. Hence N is the join of k conjugate sets, i.e. a conjugate class.

14. Prove Th. 3.2.5.6.

PROOF:

Since every element of KN is of the form kn for every $k \in K$ and $n \in N$, and since N is a normal subgroup, it immediately follows that $kN = Nk$, i.e. $kn = nk'$ for some $k' \in K$. Hence $kn \in NK$ which implies $KN \subseteq NK$. Similarly, $NK \subseteq KN$. Hence $KN = NK$.

15. Prove Th. 3.2.5.7.

PROOF:

Since any subgroup of a group G is a complex of G , it is evident, by Th. 3.2.5.6, that $SN = NS$.

If $x, y \in SN$, then $x = s_1 n_1$ and $y = s_2 n_2$, where $s_1, s_2 \in S$ and $n_1, n_2 \in N$. Moreover, $xy^{-1} = (s_1 n_1)(s_2 n_2)^{-1} = s_1 n_1 n_2^{-1} s_2^{-1} = s_1 n_3 s_2^{-1}$ where $n_1 n_2^{-1} = n_3 \in N$. Likewise, $n_3 s_2^{-1} = s_3 n_3$, where $s_2^{-1} = s_3 \in S$, since $SN = NS$. Hence $xy^{-1} = s_1 n_3 s_2^{-1} = s_1 n_3 s_3 = s_1 s_3 n_3 = (s_1 s_3) n_3 \in SN$, and conversely, which, by Th. 3.2.1.2, proves $SN = NS$ to be a normal subgroup of G .

16. Prove Th. 3.2.5.8.

PROOF:

Since $a, b \in K$ implies, by hypothesis, $a \rightarrow e'$ and $b \rightarrow e'$, it follows from Th. 3.1.3.3 that $a^{-1} \rightarrow e'^{-1} = e'$ and $a^{-1}b \rightarrow e'e' = e'$, which implies $a^{-1}b \in K$, and conversely. Hence K is a subgroup of G' .

Moreover, again by Th. 3.1.3.3, $g \in G \rightarrow g' \in G'$ implies $g^{-1} \rightarrow g'^{-1}$, which in turn implies $g^{-1}kg \rightarrow g'^{-1}kg' \rightarrow g'^{-1}e'g' = e'$, for every $k \in K \rightarrow e'$. But, then, $g^{-1}kg \in K$. Hence $g^{-1}Kg \subseteq K$ and, by Prob. 9, K is a normal subgroup of G .

17. Prove Th. 3.2.5.10.

PROOF:

If $g_1 \rightarrow g'$ and $g_2 \rightarrow g'$, where $g_1, g_2 \in G$ and $g' \in G$, then $g_1 g_2^{-1} \rightarrow e$ and $g_1 g_2^{-1} \in S$, i.e. $g_1 \in S g_2$, which proves that g_1 and g_2 are in the same coset of S .

Conversely, if $g_1 \in S g_2$, then $g_1 = s g_2$, where $s \in S$; and if $g_2 \rightarrow g'$, then $g_1 \rightarrow g'$ since $s \rightarrow e$, proving that g_1 and g_2 have the same homomorphic image in G' , completing the proof.

Note. Since this theorem is an immediate result of Th. 3.2.5.8, the former often appears as a part of the latter, i.e. the First Theorem on Homomorphism.

18. Prove Th. 3.2.5.11.

PROOF:

Since, by Th. 2.2.5.10, $xS = Sx$ for every $x \in G$, it follows that $(xS)(yS) = xSyS = xySS$. But $SS = S$ for any subgroup S of a group G , since, by Df. 3.2.2.1, SS is the complex of all elements of the form ss , for any $s \in S$, which then, by the closure property of S , yields all distinct elements of S . Hence $(xS)(yS) = xyS$.

19. Prove Th. 3.2.5.12.

PROOF:

(i) If $a, b \in G_1$, then $a \rightarrow a'$ and $b \rightarrow b'$, $a', b' \in H(G_1)$, under $H: G \rightarrow G'$ and also $a^{-1}b \in G_1 \rightarrow a'^{-1}b' \in H(G_1)$, proving that $H(G_1)$ is a subgroup of $H(G)$.

If $G_1 = G$, then $H(G_1) = H(G)$ is a subgroup of G' , i.e. a subgroup of the homomorph of G itself.

(ii) Since $g \in G \rightarrow g' \in G'$ and $h \in G_1 \rightarrow h' \in H(G_1)$, it follows that $g^{-1}hg \rightarrow g'^{-1}h'g'$, which implies $g'^{-1}h'g' \in H(G_1)$ since $g^{-1}hg \in G_1$. Hence $g'^{-1}H(G_1)g' \subseteq H(G_1)$, proving that, by Prob. 9 above, $H(G_1)$ is a normal subgroup of $H(G)$.

20. Given the multiplicative group R^* of all real numbers and $x \in R^*$, find homomorphisms and their kernels in the following correspondences: (i) $x \rightarrow |x|$, (ii) $x \rightarrow 1/x^2$.**Solution:**

(i) Since there generally exists a two-one correspondence: $x, -x \rightarrow |x|$, where $x \rightarrow |x|$ and $y \rightarrow |y|$ imply $xy \rightarrow |xy| = |x||y|$, this is a homomorphism whose kernel is a complex $\{1, -1\}$ which does form a normal subgroup.

(ii) A homomorphism is defined here likewise, since $x \rightarrow 1/x^2$ and $y \rightarrow 1/y^2$ do imply $xy \rightarrow 1/(xy)^2 = (1/x^2)(1/y^2)$ in a two-one correspondence; the kernel is here again $\{1, -1\}$.

21. In a homomorphism H of a group G onto or into a group G' , the complex C of all elements of G corresponding to the elements of a complex C' , $c' \in G'$, is the join of all cosets of the kernel K of H . If C' is a subgroup of G' , then C is a subgroup of G , and if C' is a normal subgroup of G' , so is C of G .

PROOF:

- (i) Since $c \in C \rightarrow c' \in C'$ under H , and since $k \in K \rightarrow e' \in G'$, it follows that $ck \in cK \rightarrow e'c' = c' \in C$. Hence $ck \in C$, which implies $cK \subseteq C$, proving that C is the join of all cosets cK , $c \in C$.
- (ii) If C' is a subgroup of G' , then $a \in C \rightarrow a' \in C'$ and $b \in C \rightarrow b' \in C'$ under H , which implies $a^{-1}b \in C \rightarrow a'^{-1}b' \in C'$, proving that C must be a subgroup of G if C' is a subgroup of G' .
- (iii) Likewise, $g \in G \rightarrow g' \in G'$ and $c \in C \rightarrow c' \in C'$ imply $g^{-1}cg \in C \rightarrow g'^{-1}c'g' \in C'$, which in turn implies $g^{-1}Cg \subseteq C$ since $g^{-1}cg \in C$. Hence C is a normal subgroup of G if C' is a normal subgroup of G' .

22. If G is an Abelian group and G_n is a set which consists of all g^i , where $g \in G$, $i = 1, 2, \dots, n$, then there exists a homomorphism H of G onto G_n where the kernel K of H is the set of every element k such that $k^n = e$.

PROOF:

The homomorphism H does exist since $a \in G \rightarrow a^n \in G_n$ and $b \in G \rightarrow b^n \in G_n$ imply $ab \in G \rightarrow (ab)^n = a^n b^n \in G_n$ where the kernel K of H cannot but be the set of k such that $k^n = e$.

23. The quaternion group $Q = \{e, a, a^2, a^3, b, ab, a^2b, a^3b\}$, defined by $e = a^4$, $a = b = c$, $a = bc$, $b = ca$, $c = ab$, is a non-commutative group of order 8. Verify that all subgroups of Q are normal.

PROOF:

Q has four proper subgroups: one, of order 2, is $\{a^2\}$, and the other three, of order 4, are $\{a\}$, $\{b\}$, and $\{ab\}$, which are indeed all normal, as can be easily verified by computation with the multiplication table at the end of §3.2.4.

Note. The eight elements of Q , $e, a, a^2, a^3, b, ab, a^2b, a^3b$ may be transformed into $1, i, -1, -i, j, k, -j, -k$, respectively, where the quaternions have the following properties:

$$i^2 = j^2 = k^2 = -1, \quad ij = k = -ji, \quad jk = i = -kj, \quad ki = j = -ik$$

24. If an Abelian group G is simple, the order of G is a prime.

PROOF:

All subgroups of $G \neq \{e\}$ are also Abelian and, consequently, normal. Hence, by Th. 3.2.2.12, the order of G is of a prime.

*§3.2.6 Quotient Groups

Th. 3.2.6.1 The cosets of any normal subgroups G_i of a group G form a group under complex multiplication (cf. Df. 3.2.2.1, and Prob. 1 below).

Df. 3.2.6.2 The group of cosets of G_i in Th. 3.2.6.1 is called the *quotient group* of G by G_i , or the *factor group* of G_i in G , denoted by G/G_i .

Example:

In the example of Th. 3.2.5.5, V_4 is a normal subgroup of A_4 and necessarily also of S_4 , which then has the following decomposition, considering $G \leftrightarrow S_4$ and $G_i \leftrightarrow V_4$ in Df. 3.2.5.1:

$$\begin{aligned}
S_4 &= (1)V_4 \cup (12)V_4 \cup (13)V_4 \cup (14)V_4 \cup (123)V_4 \cup (132)V_4 \\
&= ((1), (12)(34), (13)(24), (14)(23)) \cup ((12), (34), (1324), (1423)) \\
&\quad \cup ((13), (1234), (24), (1423)) \cup ((14), (1243), (1342), (23)) \\
&\quad \cup ((123), (134), (243), (142)) \cup ((132), (234), (124), (142)) \\
&= V_4(1) \cup V_4(12) \cup V_4(13) \cup V_4(14) \cup V_4(123) \cup V_4(132)
\end{aligned}$$

S_4 , then, decomposed into six cosets of V_4 , yields the quotient group of S_4 by V_4 , denoted by

$$S_4/V_4: (1)V_4, (12)V_4, (13)V_4, (14)V_4, (123)V_4, (132)V_4$$

Df. 3.2.6.3 If G and G_i are of order m and n respectively, then the quotient group G/G_i is of order m/n , i.e. the index of G_i in G .

Example:

S_4/V_4 is of order $24/4 = 6$, as is quite obvious in the above example. Not so obvious, perhaps, but more important in the same example is what is implicit there, viz. an isomorphism between S_4/V_4 and S_3 , both of which are of order 6 (cf. Prob. 2).

Df. 3.2.6.4 If the quotient group G/G_i is considered in terms of addition (instead of the usual multiplication), it is then called the *difference group* of G by G_i , denoted by $G - G_i$.

Example:

If G represents the set I of all integers, which is a module, i.e. an Abelian additive group, then the complex $G_i = \{n\}$, i.e. the set of all multiples of positive integer n , is by definition a normal subgroup. The difference group $I - \{n\}$ then consists of the cosets of the form: $a + \{n\}$, $a \in I$.

Th. 3.2.6.5 The difference group in I , $I - \{n\}$, is isomorphic to I itself. (Cf. Prob. 7.)

Df. 3.2.6.6 The coset of the form $a + \{n\}$, $a \in I$, in the example of Df. 3.2.6.4 is called the *residue class modulo n* , and the difference group in Th. 3.2.6.5 is called the *additive group of residue class modulo n* . (Cf. Prob. 4, 8-10.)

Example:

For $n = 4$, I is decomposed into the following 4 cosets:

$$\begin{aligned}
\dots -12, -8, -4, 0, 4, 8, 12, \dots &= \{0\} = 4k = M_0 \\
\dots -11, -7, -3, 1, 5, 9, 13, \dots &= \{1\} = 4k+1 = M_1 \\
\dots -10, -6, -2, 2, 6, 10, 14, \dots &= \{2\} = 4k+2 = M_2 \\
\dots -9, -5, -1, 3, 7, 11, 15, \dots &= \{3\} = 4k+3 = M_3
\end{aligned}$$

and the class \mathbf{M} of four sets M_0, M_1, M_2, M_3 is the additive group of residue class modulo 4, which, under addition, satisfies the multiplication table on the right.

Note. The subset $\{0, 1, 2, 3\}$ constitutes a set of *representatives* (cf. Df. 2.1.15 and Df. 3.2.2.7) for the class \mathbf{M} .

+	M_0	M_1	M_2	M_3
M_0	M_0	M_1	M_2	M_3
M_1	M_1	M_2	M_3	M_0
M_2	M_2	M_3	M_0	M_1
M_3	M_3	M_0	M_1	M_2

Df. 3.2.6.6 has an alternative form as follows:

Df. 3.2.6.6a If a and b are integers such that the difference $a - b$ is divisible by m , a is said to be *congruent to b modulo m* , denoted by $a \equiv b \pmod{m}$.

Example:

$$17 \equiv 7 \pmod{5}, \text{ since } 17 - 7 \text{ is divisible by } 5.$$

Th. 3.2.6.7 Congruence is an equivalence relation, i.e.,

- (i) $a \equiv a \pmod{m}$.
- (ii) $a \equiv b \pmod{m}$ implies $b \equiv a \pmod{m}$.
- (iii) $a \equiv b \pmod{m}$ and $b \equiv c \pmod{m}$ imply $a \equiv c \pmod{m}$. (Cf. Prob. 12.)

Th. 3.2.6.8 $a \equiv b \pmod{m}$ and $c \equiv d \pmod{m}$ imply

- (i) $a+c \equiv b+d \pmod{m}$ and $a-c \equiv b-d \pmod{m}$.
- (ii) $ac \equiv bd \pmod{m}$. (Cf. Prob. 13.)

Th. 3.2.6.9 $a \equiv b \pmod{m}$ implies $ac \equiv bc \pmod{cm}$. (Cf. Prob. 15.)

Df. 3.2.6.10 The *irreducible (residue) class modulo p* , a prime, consists of $(p-1)$ residue sets modulo p , i.e. $\{1\}, \{2\}, \dots, \{p-1\}$, which contain no integer divisible by p .

Example:

If $p = 5$, then

$$\begin{aligned}\{0\} &= \{\dots -15, -10, -5, 0, 5, 10, 15, \dots\} \\ \{1\} &= \{\dots -14, -9, -4, 1, 6, 11, 16, \dots\} \\ \{2\} &= \{\dots -13, -8, -3, 2, 7, 12, 17, \dots\} \\ \{3\} &= \{\dots -12, -7, -2, 3, 8, 13, 18, \dots\} \\ \{4\} &= \{\dots -11, -6, -1, 4, 9, 14, 19, \dots\}\end{aligned}$$

of which the set $\{0\}$ alone contains the integers divisible by 5 and as such does not belong to the irreducible residue class modulo 5.

Th. 3.2.6.11 The irreducible class modulo p forms a multiplicative group. (Cf. Prob. 16.)

Th. 3.2.6.12 The irreducible class modulo p has a subgroup which consists of $\{1\}$ and $\{p-1\}$. (Cf. Prob. 17.)

Th. 3.2.6.13 (by Fermat). If a is an integer and p a prime, then $a^p \equiv a \pmod{p}$. (Cf. Prob. 18.)

Stated otherwise: If an integer a is not divisible by a prime p , then $a^{p-1} \equiv 1 \pmod{p}$.

Th. 3.2.6.14 If an element a of a group G has a prime order, then $b^{-1}ab = a^n$, $b \in G$, implies that a and b^{p-1} are permutable. (Cf. Prob. 19.)

Th. 3.2.6.15 If N is a normal subgroup of G , then the mapping of G onto G/N : $g \rightarrow gN$, $g \in G$, is a homomorphism. (Cf. Prob. 20.)

Df. 3.2.6.16 The homomorphism: $G \rightarrow G/N$, defined by Th. 3.2.6.15, is called the *natural homomorphism* of G onto G/N .

Th. 3.2.6.17 (Second Theorem on Homomorphism). If a homomorphism $H: G \rightarrow Q$, where N is a normal subgroup of G and $Q = G/N$, is obtained by the correspondence: $g \rightarrow aK$ where $g \in aK$, then K is the kernel of H . (Cf. Prob. 21.)

Th. 3.2.6.18 (First Theorem on Isomorphism). If there exists a homomorphism $H: G \rightarrow G'$ whose kernel is K , then G' is isomorphic to $Q = G/K$. (Cf. Prob. 22.)

Th. 3.2.6.18 is sometimes called the Third Theorem on Homomorphism, making Th. 3.2.7.3 the First Theorem on Isomorphism.

Solved Problems

1. Prove Th. 3.2.6.1.

PROOF:

G1: Closure. Let $a, b \in G$, and $ag_1 \in aG_i$, $bg_2 \in bG_i$; then, since G_i is normal,

$$(ag_1)(bg_2) = abb^{-1}g_1bg_2 = ab(b^{-1}g_1bg_2) = ab(g_1g_2) \in abG_i$$

Hence the product of any two elements (e.g. aG_i, bG_i) of a coset of G_i is uniquely determined (i.e. abG_i) to be a distinct element of the same coset. (Cf. Th. 3.2.5.11.)

G2: Associative Law. Let $a, b, c \in G$; then, by **G1**,

$$aG_i((bG_i)(cG_i)) = aG_i(bcG_i) = abcG_i = (abG_i)cG_i = (aG_i)(bG_i)cG_i = ((aG_i)(bG_i))cG_i$$

G3: Identity. Since $eG_iaG_i = eaG_i = aG_i$ and $aG_ieG_i = aeG_i = aG_i$, the identity of the coset is $eG_i = G_i$.

G4: Inverse. Since $aG_ia^{-1}G_i = aa^{-1}G_i = G_i$ and $a^{-1}G_iaG_i = a^{-1}aG_i = G_i$, $a^{-1}G_i$ is the inverse of aG_i .

Hence the cosets of any normal subgroup form a group.

2. Set up an isomorphism between S_4/V_4 and S_3 .

Solution:

The desired isomorphism can be established as follows:

$$\begin{array}{ll} V_4(1) & \leftrightarrow (1) \\ V_4(13) & \leftrightarrow (13) \\ V_4(123) & \leftrightarrow (123) \end{array} \quad \begin{array}{ll} V_4(12) & \leftrightarrow (23) \\ V_4(14) & \leftrightarrow (12) \\ V_4(132) & \leftrightarrow (132) \end{array}$$

as can be readily justified by computing with the multiplication table of S_3 ; e.g.,

$$(V_4(13))(V_4(14)) = V_4(123) \leftrightarrow (13)(12) = (123), \text{ etc.}$$

3. Verify that there exists an isomorphism between the quotient group S_n/A_n of the alternating group A_n in the symmetric group S_n and a group of order 2 which is defined by the multiplication table on the right.

\times	0	1
0	0	1
1	1	0

PROOF:

Let p be an odd permutation; then S_n/A_n consists of A_n and pA_n (cf. Th. 3.2.5.5), and an isomorphism is established as follows

$$A_n \leftrightarrow 0 \quad \text{and} \quad pA_n \leftrightarrow 1$$

which is justified by the following 1-1 correspondences:

$$\begin{array}{ll} A_n A_n = A_n & \leftrightarrow 00 = 0 \\ pA_n A_n = pA_n & \leftrightarrow 10 = 1 \end{array} \quad \begin{array}{ll} A_n pA_n = pA_n & \leftrightarrow 01 = 1 \\ pA_n pA_n = A_n & \leftrightarrow 11 = 0 \end{array}$$

Note. In general, the factor group G/N of a normal subgroup N of index n in G is of order n . Note, also, that the multiplication table above can be used for the additive group $\{0,1\}$ of residue class modulo 2 or the additive (but not multiplicative) group $\{E, O\}$, E representing any even number and O any odd number.

4. Given a normal subgroup N of a group G , the factor group G/N is cyclic if the index in G of N is a prime.

PROOF:

Since the order of G/N is a prime $p = (G:N)$, it follows at once from Th. 3.2.2.12 that G/N is cyclic.

5. For any subgroup S of an infinite cyclic group G , G/S is a finite cyclic group of order n if $(G:S) = n$.

PROOF:

Let a generator of G be g and an element of S whose exponent is the smallest be g^m ; then S is a cyclic subgroup generated by g^m . The cosets of S are, then: $S, gS, g^2S, \dots, g^{m-1}S$. Hence $n = m$, and G/S is a cyclic subgroup of order n , generated by gS .

6. Given a normal subgroup N of a group G , any two subgroups H and K of G are conjugate to each other in G if the subgroups $H' = H/N$ and $K' = K/N$ of $G' = G/N$ are conjugate to each other in G' , and conversely.

PROOF:

Let $g' = gN$, $g \in G$, and $g'^{-1}H'g' = K'$; then, for every $h \in H$, $(gN)^{-1}(hN)(gN) = g^{-1}hgN \in K$. Hence $g^{-1}hg \in K$, i.e. $g^{-1}Hg \subseteq K$.

Likewise, since $gkg^{-1} \in H$, it follows that $gKg^{-1} \subseteq H$; hence $K \subseteq g^{-1}Hg$. Hence, altogether, $g^{-1}Hg = K$.

The converse can be similarly proved.

7. Prove Th. 3.2.6.5.

PROOF:

Since the difference group: $I - \{n\}$ consists of the cosets of the form $a + \{n\}$, $a \in I$, let $x \in I \leftrightarrow x + \{n\}$ and $y \in I \leftrightarrow y + \{n\}$, which together imply $x + y \in I \leftrightarrow (x + \{n\}) + (y + \{n\}) = (x + y) + \{n\} \in I - \{n\}$, completing the proof.

8. Verify an isomorphism between the additive group $A = \{0, 1, 2, 3\}$ of residue class modulo 4 and the cyclic group C of order 4.

PROOF:

The general 1-1 correspondence between A and C is $n \leftrightarrow a^n$, and individually:

$$0 \leftrightarrow a^0 = e, \quad 1 \leftrightarrow a, \quad 2 \leftrightarrow a^2, \quad 3 \leftrightarrow a^3$$

The multiplication table of C is then the same as that of A .

Note. The table in Df. 3.2.6.6 is the same as that of the arithmetic of changing tires (cf. §3.1.2, Prob. 1); so is also the multiplication table of the subgroup $\{0, 1, 2, 3\}$ for the symmetries of rotations of the dihedral group, (cf. §3.1.2, Prob. 5).

Note also that V_4 is an exception.

9. The cyclic group C of order n is isomorphic to the additive group A of residue class modulo n .

PROOF:

Since C consists of $e, a, a^2, \dots, a^{n-1}$ and A of $\{0\}, \{1\}, \{2\}, \dots, \{n-1\}$, let $a^i \leftrightarrow \{i\} = A_i$. The correspondence is 1-1 and also distinct; for, if $a^i = a^j$, then $i \equiv j \pmod{n}$, which in turn implies $A_i = A_j$, and conversely, if $A_i = A_j$, then $i \equiv j \pmod{n}$ which implies an integer k such that $i - j = kn$ and, consequently, $a^i = a^{j+kn} = a^j(a^n)^k = a^j e^k = a^j$.

Furthermore, $a^i \leftrightarrow A_i$ and $a^j \leftrightarrow A_j$, where $i \not\equiv j \pmod{n}$, imply $a^i a^j = a^{i+j} \leftrightarrow A_{i+j} = A_i + A_j$, completing the proof.

Note. Prob. 9 is, of course, the result of generalizing Prob. 8.

10. Any factor group of a cyclic group is cyclic.

PROOF:

Let H be a homomorphic mapping of a cyclic group $G = \{a\}$ of order n onto a cyclic group G' of order m . Since every element of G is a power of a , $H(a) = b$ implies that any element b of G' is also a power of b , i.e. $G' = \{b\}$. Hence G/G' is a cyclic group of order n/m .

Note. If G is an infinite cyclic group isomorphic to the additive group I of all integers, a homomorphism H of G into a cyclic group G' of order m is obtained if $i \in I$ is assigned to the element b^i as its homomorphic image. Thus i and j are mapped onto the same element of G' , iff $i \equiv j \pmod{m}$. In I , then, the residue class modulo n corresponds to G .

11. The residue class modulo m is a quotient group of order m .

PROOF:

Let G be the additive group of all positive integers, which is Abelian, and N be a subgroup of G which consists of all multiples of m . If N_i is to designate a coset of N to which the integer i belongs, then N has $(m-1)$ cosets in G : N_0, N_1, \dots, N_{m-1} , where $N_0 = N$ and $N_i = N_j$ iff $i \equiv j \pmod{m}$, which by definition is the residue class modulo m . But, G being an Abelian group, N is also Abelian and, consequently, normal. Hence, by Df. 3.2.6.2, G/N yields a quotient group of order m , which is indeed the Abelian additive group of the residue class modulo m .

Note. **G1:** $N_{i+j} = N_i + N_j$; **G2:** $N_i + (N_j + N_k) = N_i + (j+k) = N(i+j) + k = (N_i + N_j) + N_k$; **G3:** N_0 ; **G4:** $N_{-i} = -N_i$.

12. Prove Th. 3.2.6.7.

PROOF:

$$(i) \quad a - a = 0 = 0 \cdot m.$$

$$(ii) \quad \text{If } a - b = mq, q \in I, \text{ then } b - a = m(-q), -q \in I.$$

$$(iii) \quad \text{If } a - b = mq_1, q_1 \in I, \text{ and } b - c = mq_2, q_2 \in I, \text{ then}$$

$$a - c = (a - b) + (b - c) = mq_1 + mq_2 = m(q_1 + q_2), \quad q_1 + q_2 \in I$$

13. Prove Th. 3.2.6.8.

PROOF:

If $a - b = mq_1, q_1 \in I$, and $c - d = mq_2, q_2 \in I$, then

$$(i) \quad (a + c) - (b + d) = m(q_1 + q_2), \quad q_1 + q_2 \in I \quad \text{and}$$

$$(a - c) - (b - d) = m(q_1 - q_2), \quad q_1 - q_2 \in I;$$

$$(ii) \quad ac - bd = a(c - d) + d(a - b) = m(aq_2 + dq_1), \quad aq_2 + dq_1 \in I.$$

14. Generalize Th. 3.2.6.8.

PROOF:

If $a_1 \equiv b_1 \pmod{m}, a_2 \equiv b_2 \pmod{m}, \dots, a_n \equiv b_n \pmod{m}$, let $a_1 - b_1 = mq_1, a_2 - b_2 = mq_2, \dots, a_n - b_n = mq_n, q_1, q_2, \dots, q_n \in I$; then:

$$(i) \quad (a_1 + a_2 + \dots + a_n) - (b_1 + b_2 + \dots + b_n) = m(q_1 + q_2 + \dots + q_n), \quad q_1 + q_2 + \dots + q_n \in I$$

$$\text{and} \quad (a_1 - a_2 - \dots - a_n) - (b_1 - b_2 - \dots - b_n) = m(q_1 - q_2 - \dots - q_n), \quad q_1 - q_2 - \dots - q_n \in I$$

proving that

$$a_1 + a_2 + \dots + a_n \equiv b_1 + b_2 + \dots + b_n \pmod{m}$$

and

$$a_1 - a_2 - \dots - a_n \equiv b_1 - b_2 - \dots - b_n \pmod{m}$$

(ii) Since the case for $n = 2$ has already been proved (cf. Prob. 13ii), assume, for $n = k$,

$$a_1 a_2 \dots a_k - b_1 b_2 \dots b_k = mq_s, \quad q_s \in I$$

Then, for $n = k + 1$,

$$\begin{aligned} a_1 a_2 \dots a_k a_{k+1} - b_1 b_2 \dots b_k b_{k+1} &= a_1 a_2 \dots a_k (a_{k+1} - b_{k+1}) + b_{k+1} (a_1 a_2 \dots a_k - b_1 b_2 \dots b_k) \\ &= m(a_1 a_2 \dots a_k q_{k+1} + b_{k+1} q_s), \quad a_1 a_2 \dots a_k q_{k+1} + b_{k+1} q_s \in I \end{aligned}$$

Hence $a_1 a_2 \dots a_n \equiv b_1 b_2 \dots b_n \pmod{m}$.

Note. If $a_1 = a_2 = \dots = a_n$ and $b_1 = b_2 = \dots = b_n$ in (ii), then $a^n \equiv b^n \pmod{m}$.

15. Prove Th. 3.2.6.9.

PROOF:

Since $a - b = mq, q \in I$, it follows immediately that $ac - bc = c(a - b) = cmq$.

Note. If $a \equiv b \pmod{m}$, then also $ac \equiv bc \pmod{m}$, but not conversely; e.g. $10 \equiv 2 \pmod{8}$ does not imply $5 \equiv 1 \pmod{8}$. In fact, $ac \equiv bc \pmod{m}$ implies only $a \equiv b \pmod{m/(c,m)}$ where (c,m) denotes a common divisor of c and m (cf. §4.1.2.3, Prob. 39).

16. Prove Th. 3.2.6.11.

PROOF:

Let $p=5$, for the sake of simplicity, as is outlined in the example of Df. 3.2.6.10; then the irreducible class \mathbf{M} modulo 5 consists of four sets: $M_1 = \{1\}$, $M_2 = \{2\}$, $M_3 = \{3\}$, $M_4 = \{4\}$, their representatives being 1, 2, 3, 4. Then, as the multiplication table on the right exemplifies,

G1: $M_a M_b = M_{ab}$, $a, b = 1, 2, 3, 4$; and if $M_a = M_b$ and $M_c = M_d$, then, by Th. 3.2.6.8, $M_{ab} = M_{cd}$.

G2: $M_a(M_b M_c) = M_a M_{bc} = M_{abc} = M_{ab} M_c = (M_a M_b) M_c$

G3: $M_a M_1 = M_1 M_a = M_a$

G4: $M_1 M_1 = M_2 M_3 = M_3 M_2 = M_4 M_4 = M_1$

	1	2	3	4
1	1	2	3	4
2	2	4	1	3
3	3	1	4	2
4	4	3	2	1

17. Prove Th. 3.2.6.12.

PROOF:

Since $M_1 M_1 = M_1$, and $M_{p-1} M_{p-1} = M_p M_{-1} M_p M_{-1} = M_{-1} M_{-1} = M_1 M_1 = M_1$, and $M_1 M_{p-1} = M_{p-1} M_1 = M_{p-1}$, it follows that M_1 and M_{p-1} put themselves in a subgroup by themselves.

18. Prove Th. 3.2.6.13.

PROOF:

The order of the multiplicative group \mathbf{M} of an irreducible class modulo p is $p-1$, as has already been examined in Df. 3.2.6.10 and Th. 3.2.6.11. Since, by Th. 3.2.2.10, the order of any element M_a , $a \neq 0$, of \mathbf{M} is a divisor of the order of \mathbf{M} , it follows that $(M_a)^{p-1} = M_1$. (E.g., $1^4 \equiv 2^4 \equiv 3^4 \equiv 4^4 \equiv 1 \pmod{5}$.) But, obviously, $a^{p-1} \in (M_a)^{p-1}$. Hence $a^{p-1} \equiv 1 \pmod{p}$, i.e. $a^p \equiv a \pmod{p}$.

If $a = 0$, i.e. $a \in \{0\}$, then a must be divisible by p , i.e. $a \equiv 0 \pmod{p}$, and the theorem evidently holds.

(This theorem reveals, e.g., $(10)^{10} \equiv 1 \pmod{11}$, $(28)^{30} \equiv (29)^{30} \equiv (30)^{30} \equiv 1 \pmod{31}$, etc.)

19. Prove Th. 3.2.6.14.

PROOF:

Since, for $a^n = e$, $a = be b^{-1} = bb^{-1} = e$, which is contradicting the hypothesis of a being the generator of G , n must not be 0; n , then, represents an irreducible class modulo p : $1, 2, \dots, p-1$.

Now, $b^{-1}ab^1 = a^n$ implies $b^{-2}ab^2 = b^{-1}b^{-1}abbb = b^{-1}a^n b = (b^{-1}ab)^n = (a^n)^n = a^{n^2}$. Suppose $b^{-k}ab^k = a^{n^k}$; then $b^{-(k+1)}ab^{(k+1)} = b^{-1}(b^{-k}ab^k)b = b^{-1}a^{n^k}b = a^{n^{(k+1)}}$. Hence $b^{-(p-1)}ab^{p-1} = a^{n^{(p-1)}}$.

But then, by Th. 3.2.6.13, there exists an integer q such that

$$n^{p-1} = 1 + pq, \quad \text{and} \quad b^{-(p-1)}ab^{(p-1)}a^{n^{(p-1)}} = a^{(1+pq)} = a(a^p)^q = ae^q = a$$

Hence, $ab^{p-1} = b^{p-1}a$.

20. Prove Th. 3.2.6.15.

PROOF:

By Th. 3.2.5.10, $a_1 \in G \rightarrow a_1 N \in aN$ and $a_2 \in G \rightarrow a_2 N \in aN$ imply $a_1 a_2 \in G \rightarrow (a_1 N)(a_2 N) = a_1 a_2 N \in aN$.

21. Prove Th. 3.2.6.17.

PROOF:

Let $g_1 \in a_i K$ and $g_2 \in a_j K$ in the mapping $g \rightarrow aK$; then, by Th. 3.2.5.11, $g_1 g_2 \in a_k K$ where $a_i a_j \in a_k K$. Hence $g_1 g_2 \rightarrow a_k K = a_i K a_j K$, proving the mapping of G onto Q to be H , where $g \rightarrow e$ iff $g \in K$ in G , since K is the identity of Q . Hence K is the kernel of H .

22. Prove Th. 3.2.6.18.**PROOF:**

It follows directly from Th. 3.2.5.10 that every $g \in G$ in the same coset of K has the same image in G' , and consequently that the correspondence $x' \leftrightarrow xK$ is 1-1. But, since $x \rightarrow x'$ and $y \rightarrow y'$ imply $xy \rightarrow x'y'$ where $xy \in xyK$, it also follows that $x'y' \leftrightarrow xyK = xKyK$. Hence the correspondence $x' \leftrightarrow xK$ is 1-1 between G' and Q .

Note. Th. 3.2.6.18 may be given in an alternative form, which as such needs a longer proof; viz. if K is a complex $\{k\}$ of a group G which is mapped onto a group G' under a homomorphism H , then $H: k \in K \rightarrow e' \in G'$ implies that K is a normal subgroup of G under H , hence the kernel of H , and that $G/K \rightarrow G'$ is an isomorphism.

Proof. The complex K is a group, since $a, b, c \in K$ implies **G1:** $H(ab) = H(a)H(b) = e'e' = e'$; **G2:** $H(a(bc)) = H(a)(H(b)H(c)) = (H(a)H(b))H(c) = H((ab)c)$; **G3:** $H(e) = e'$; **G4:** $e' = H(e) = H(aa^{-1}) = H(a)H(a^{-1}) = e'H(a^{-1}) = H(a^{-1})$.

K is also normal; it is in fact the kernel of H , since $h = g^{-1}kg$ for every $g \in G$ and every $k \in K$ implies $h \in K$ (since $H(h) = H(g^{-1}kg) = H(g^{-1})H(k)H(g) = H(g^{-1})e'H(g) = H(g^{-1}g) = H(e) = e'$).

Furthermore, since every $g' \in G'$ is of the form $H(a)$ for some $a \in G$, and since every element in G/K can be written as aK , the mapping of G' onto G/K is 1-1; so is the mapping of G/K onto G' , since $H(a) = H(b)$ iff $aK = bK$. Hence the mapping is 1-1, and also isomorphic, since $T(H(a)) = aK$ defines the isomorphic mapping of G' into G/K and $T(H(a)H(b)) = T(H(ab)) = (ab)K = (aK)(bK) = T(H(a))T(H(b))$.

*§3.2.7 Composition Series and Direct Products

Df. 3.2.7.1 A normal subgroup N of a group G is *maximal* if it is not properly contained by any proper normal subgroup of G .

Stated otherwise: if N is a maximal normal subgroup of G , then there exists no normal subgroup N' such that $N \subset N' \subset G$.

Th. 3.2.7.2 A normal subgroup N of G is maximal in G iff G/N is simple. (Cf. Df. 3.2.5.2a and Prob. 2.)

Th. 3.2.7.3 (Second Theorem on Isomorphism). If N is a normal subgroup, and H any subgroup, of a group G , then the meet M of N and H is normal and the correspondence $NH/N \leftrightarrow H/M$ is an isomorphism. (Cf. Prob. 3.)

Df. 3.2.7.4 A series of subgroups of a group G :

$$\{e\} = S_0 \subset S_1 \subset \dots \subset S_n = G$$

where S_i/S_{i-1} (called a *composition-quotient group*) is simple, is called a *composition series*. The order of a composition quotient group is called the *composition index*.

It is evident that S_{i-1} is a maximal normal subgroup of S_i . The following theorem is also evident.

Th. 3.2.7.5 There exists at least one composition series with respect to a group G . (Cf. Prob. 5.)

Example:

$\{(1)\} \subset \{(12)(34)\} \subset V_4 \subset A_4 \subset S_4$ (cf. the example in Df. 3.2.5.5), where $\{(12)(34)\}$ represents three subgroups, excluding (1) of V_4 , which are all of order 2.

Hence the composition indices are: 2, 2, 3, 2.

Th. 3.2.7.6 (by Jordan-Hölder). If G is a group with two composition series C_1 and C_2 ,

$$C_1: \{e\} = S_0 \subset S_1 \subset \dots \subset S_m \subset S_{m+1} = G$$

$$C_2: \{e\} = S'_0 \subset S'_1 \subset \dots \subset S'_n \subset S'_{n+1} = G$$

then $m = n$, and there exists an isomorphism: $S_i/S_{i-1} \leftrightarrow S'_i/S'_{i-1}$. (Cf. Prob. 6).

Note the difference in the subscripts of S and S' in the isomorphism specified above; the isomorphism takes place in *some* order and not always exactly opposite to each other.

Example:

If G is a cyclic group of order 6, $\{e, a, a^2, a^3, a^4, a^5\}$, then two composition series are $e \subset S_1 = \{e, a^2, a^4\} \subset G$ and $e \subset S'_1 = \{e, a^3\} \subset G$, and

$$G/S'_1 \leftrightarrow S_1/e \quad \text{and} \quad G/S'_1 \leftrightarrow S_1/e$$

Another emphasis should be put on the fact, which is implicit in the form of the isomorphism, that each subgroup of the composition series is a maximal normal subgroup of the preceding group alone and may not be even normal in G itself or in any other subgroup but the preceding one.

Th. 3.2.7.7 A set S of ordered pairs (a, b) , $a \in A$, $b \in B$, where A and B are two groups under a binary operation

$$(a_1, b_1)(a_2, b_2) = (a_1a_2, b_1b_2), \quad \text{where } a_1, a_2 \in A, b_1, b_2 \in B$$

forms a group. (Cf. Prob. 7.)

Df. 3.2.7.8 The group S of Th. 3.2.7.7 is called the *direct product* (cf. Df. 2.2.2.3) of A and B , denoted by $A \times B$.

As is but logical, a_1a_2 and b_1b_2 in the binary operation defined above are obtained by the operative rules of A and B respectively.

Example:

If A represents the additive group of all integers and B the multiplicative group of the fourth roots of unity (cf. §3.1.3, Prob. 7), a_1a_2 is obtained by addition and b_1b_2 by multiplication.

If A and B are two subgroups of a group G , then the definition of the direct product of A and B is modified as follows:

Df. 3.2.7.8a Two subgroups A and B of a group G form a direct product if the following two conditions are satisfied:

(i) $ab = ba$ for every $a \in A$, $b \in B$.

(ii) $A \cap B = e$

In particular, if $G = A \times B$, then A and B are called the *direct factors* of the decomposition, which stipulates the third condition:

(iii) $g \in G$ implies a unique representation $g = ab$ for every $a \in A$, $b \in B$.

Further generalized, the direct product of more than two subgroups is defined as follows:

Df. 3.2.7.8b A group G is called the direct product of its subgroups G_1, G_2, \dots, G_n if the following three conditions are satisfied:

(i) G_1, G_2, \dots, G_n are all normal in G .

(ii) $G = \cup G_i, \quad i = 1, 2, \dots, n$

(iii) $G_j \cap (\cup_{i \neq j} G_i) = e, \quad j = 1, 2, \dots, n$

Df. 3.2.7.8b, as well as Df. 3.2.7.8a, may be considered a theorem (cf. Prob. 11), since it can be deduced from Th. 3.2.7.7 above and Th. 3.2.7.9 below.

Th. 3.2.7.9 If A and B are normal subgroups of a group G such that $A \cup B = G$ and $A \cap B = e$, then G is isomorphic to $A \times B$. (Cf. Prob. 10.)

Example:

$A = \{e, a\}$ and $B = \{e, b\}$ are normal subgroups of

$$G = \{e, a, b, ab\} \leftrightarrow \{(e, e), (e, a), (e, b), (a, b)\} = A \times B$$

This simple example makes it quite clear that Th. 3.2.7.9 may be stated as follows:

If A and B are subgroups of a group G such that $ab = ba$, for every $a \in A, b \in B$, and if there uniquely exists $g = ab$ for every $g \in G$, then A and B have no element in common except the identity $e \in G$, and $G \leftrightarrow A \times B$.

Stated this way, Th. 3.2.7.9 becomes logically equivalent to Df. 3.2.7.8a with three conditions.

Th. 3.2.7.10 If A is an Abelian group of order k which is a product of two relative primes i and j such that $b^i = c^j = e$, for every $b \in B, c \in C$, then B of all b and C of all c are subgroups of A and $B \times C = A$.

Generalized by induction, Th. 3.2.7.10 takes the following form, called the Basis Theorem for Finite Abelian Groups:

Th. 3.2.7.10a If G is an Abelian group of order $n = \prod p_i^{a_i}, i = 1, 2, \dots, k$, where p_i is a distinct prime, then $G = G_1 \times G_2 \times \dots \times G_k$, where each G_i is of order $p_i^{a_i}$.

Df. 3.2.7.11 The set of p_i in Th. 3.2.7.10a is called the *minimal generating system* and also, with respect to the set of G_i which p_i generates, the *basis* for G .

Example:

$V_4 = \{e, a, b, c\}$ is an Abelian group where $e = a^2 = b^2 = c^2$, $ab = ba = c$, $bc = cb = a$, $ca = ac = b$; furthermore, $e = a^0 b^0$, $a = a^1 b^0$, $b = a^0 b^1$, $c = a^1 b^1$. Hence a and b constitute a minimal generating system of, or a basis for, V_4 .

Note that a and b form a distinct basis, but not the unique basis, since, e.g., a and c constitute just as well a basis for V_4 .

Note, also, that Th. 3.2.7.10, and consequently Th. 3.2.7.10a, may not hold conversely.

Example:

$S_3 = \{(1), (12), (13), (23), (123), (132)\}$, which is definitely non-Abelian, also has a basis, or rather, bases, viz. $a = (12)$ and $b = (13)$ generate: $a^0 b^0 = (1)$, $a^1 b^0 = (12)$, $a^0 b^1 = (13)$, $a^1 b^1 a^1 = (12)(13)(12) = (23)$, $a^1 b^1 = (12)(13) = (123)$, $b^1 a^1 = (13)(12) = (132)$; this basis may be replaced by, e.g., (12) and (23) .

Solved Problems

1. If there exists a normal group N_2 of a group G such that $N_1 \subset N_2 \subset G$ when N_1 is also a normal group of G , then N_2/N_1 is a normal subgroup of G/N_1 , and conversely.

PROOF:

Since N_1 being a normal subgroup of G implies that it is also a normal subgroup of N_2 , N_2/N_1 is evidently a group; moreover, for every $g \in G$, every $a \in N_2$, every $gN_1 \subseteq G/N_1$, and every $aN_1 \subseteq N_2/N_1$,

$$(gN_1)^{-1}(aN_1)(gN_1) = N_1^{-1}g^{-1}aN_1gN_1 = g^{-1}agN_1N_1N_1 = g^{-1}agN_1 \subseteq aN_1$$

proving that N_2/N_1 is normal in G/N_1 .

Conversely, since N_1 is normal in G and, as above, $(gN_1)^{-1}(aN_1)(gN_1) = g^{-1}agN_1$, where $g^{-1}agN_1$, by hypothesis, is an element of N_2/N_1 , it follows that $g^{-1}ag$ is an element of N_2 , i.e. $g^{-1}N_2g = N_2$, proving N_2 to be normal.

2. Prove Th. 3.2.7.2.

PROOF:

Assume that N is maximal when G/N is not simple; then, letting N'/N be a normal subgroup of G/N , it follows that $N \subset N' \subset G$, as in Prob. 1 above, which contradicts the assumption. Hence N is not maximal if G/N is not simple; i.e. if N is maximal, then G/N is simple.

Conversely, assume that G/N is simple when N is not normal; then there exists a normal subgroup N' such that $N \subset N' \subset G$ and, by Prob. 1, $\{e\} = N/N \subset N'/N \subset G/N$, contradictory to the assumption. Hence N is maximal if G/N is simple.

3. Prove Th. 3.2.7.3.

PROOF:

By Th. 3.2.1.3, M is a subgroup of G and also of N which is a normal subgroup; hence, for every $g \in G$ and every $m \in M$, $g^{-1}mg$ is in N , and in H , too, since H is another maximal normal subgroup. Since both H and N , i.e. $H \cap N = M$, contain it, the meet M is a normal subgroup.

Furthermore, by Th. 3.2.5.7, $HN = NH$ is a normal group, and there exists a mapping: $h \rightarrow hN$, for every $h \in H$, which is a homomorphism f of H onto HN/N . But, since all the cosets of HN/N are given as $hnN = hN$, where $n \in N$, the mapping f is actually 1-1. Hence $h \leftrightarrow hN$, which implies $h \in N$ and, consequently, $h \in M$.

Conversely, $h \in M$ implies $hN = N$, and M is proved to be the kernel of f . Hence, by Th. 3.2.6.18, $H/M \leftrightarrow HN/N$.

4. If N and H are two distinct maximal normal subgroups in Th. 3.2.7.3, then $HN = G$, and there exist two isomorphisms: $H/M \leftrightarrow G/N$ and $N/M \leftrightarrow G/H$.

PROOF:

- (i) Since H and N are distinct, there must exist an element, say, $h \in H$ which does not belong to N ; hence $eh = h$ is an element of HN which is not contained in N , i.e. $N \subset HN$.

Now, assume that $HN \neq G$; then $N \subset HN \subset G$ (since HN is also a normal subgroup, of course, by Th. 3.2.5.7). But, by hypothesis, H is maximal and is now proved to be properly contained in HN . Hence it must be the case that $HN = G$.

- (ii) Since M is a subgroup of H , H has a left-decomposition, represented by a class equation C :

$$H = M \cup a_1M \cup \dots \cup a_kM$$

which clearly implies

$$HN = MN \cup a_1MN \cup \dots \cup a_kMN$$

since $hn \in HN$ implies $hn \in a_iMN$, i.e. $HN \subseteq MN \cup \dots \cup a_kMN$, and yet $HN \supseteq MN \cup \dots \cup a_kMN$, because $HN = G$, by Prob. 4 above.

Moreover, since $M \subseteq N$ implies $NM = N$, there follows a class equation C' :

$$G = N \cup a_1N \cup \dots \cup a_kN$$

Assume that C' is not a proper decomposition of G , containing some identical cosets, e.g. $a_i N = a_j N$; then $a_i n = a_j n'$ and $a_i a_j^{-1} = n^{-1} n' = n'' \in N$, which implies $a_i a_j^{-1} \in H$, since both a_i and a_j are in H , as is explicit in C . Also, since $a_i a_j^{-1}$ is in both H and N , it is also in M , which implies $a_i M = a_j M$, contradicting the assumption of C being a left-decomposition of H in terms of distinct cosets of M in G .

Hence C' must represent a left-decomposition of G .

Since, then, $a_i M \subset H/M$ and $a_i N \subset G/N$, there exists a 1-1 correspondence $a_i M \leftrightarrow a_i N$ which is also isomorphic; for $a_i M \leftrightarrow a_i N$ and $a_j M \leftrightarrow a_j N$ imply $T(a_i M a_j M) = T(a_i a_j M) = a_i a_j N = (a_i N)(a_j N) = T(a_i M)T(a_j M)$, proving T to be an isomorphic transformation.

5. Prove Th. 3.2.7.5.

PROOF:

If G is simple, a composition series $\{e\} \subset G$ is at once obtained.

If G is not simple, then there exists a maximal normal subgroup G_1 such that $\{e\} = G_0 \subset G_1 \subset G$, and in general, if G_i is not maximal in a composition series: $\{e\} = G_0 \subset G_1 \subset \dots \subset G_i \subset G_k = G$, there always exists a maximal normal subgroup G_j such that

$$\{e\} = G_0 \subset G_1 \subset \dots \subset G_i \subset G_j \subset G_k = G$$

where k is finite, completing the proof.

6. Prove Th. 3.2.7.6.

PROOF:

The theorem is trivially true for simple groups and also for any group whose order is a prime.

Let G be of non-prime order n , which then may be factored as

$$n = p_1 p_2 \dots p_r$$

where each p_i is a prime and, representing a certain prime, may appear more than once in the factorization, i.e. r primes are not all distinct. If $r = 0$, then $G = \{e\}$, and if $r = 1$, then G is again of prime order and, G having no proper subgroups, there exists only one composition series: $\{e\} \subset G$.

Assume that the proof has been carried out likewise up to $r = k$; then, for $r = k + 1$, G is again either simple, having only one composition series as above, or not simple, in which case there exists either a unique composition series, which is trivial, or more than one composition series which are all distinct. Let two of them be

$$\begin{aligned} C_1: \quad \{e\} &= S_0 \subset S_1 \subset \dots \subset S_m \subset S_{m+1} = G \\ C_2: \quad \{e\} &= S'_0 \subset S'_1 \subset \dots \subset S'_n \subset S'_{n+1} = G \end{aligned}$$

and if $S_m = S'_n$, the theorem at once holds by the assumption of induction, since the order s_m of S_m contains k or fewer prime factors. If $S_m \neq S'_n$, i.e. they are two distinct maximal normal subgroups of G , then, by Prob. 4 above,

$$G/S_m \leftrightarrow S'_n/M_r \quad \text{and} \quad G/S'_n \leftrightarrow S_n/M_r$$

M_r denoting the meet $S_m \cap S'_n$, which is a maximal normal subgroup of S_m and S'_n , since G/S_n and G/S'_n , and consequently S_m/M_r and S'_n/M_r , are simple by Th. 3.2.7.2. Consider, then,

$$\begin{aligned} C'_1: \quad \{e\} &= M_0 \subset M_1 \subset \dots \subset M_r \subset S_n \subset G \\ C'_2: \quad \{e\} &= M_0 \subset M_1 \subset \dots \subset M_r \subset S'_n \subset G \end{aligned}$$

and there exists an isomorphism:

$$M_1/M_0, \dots, S_n/M_r, G/S_n \leftrightarrow M_1/M_0, \dots, S'_n/M_r, G/S'_n$$

except for order (i.e. at least the permutation of the first two groups in this case). But, then, by the assumption of induction,

$$\begin{aligned} M_1/M_0, \dots, S_n/M_r, G/S_n &\leftrightarrow S_1/S_0, \dots, S_n/S_{n-1}, G/S_n \\ M_1/M_0, \dots, S'_n/M_r, G/S'_n &\leftrightarrow S'_1/S'_0, \dots, S'_n/S'_{n-1}, G/S'_n \end{aligned}$$

except possibly for permutations. Hence

$$S_1/S_0, \dots, S_n/S_{n-1}, G/S_n \leftrightarrow S'_1/S'_0, \dots, S'_n/S'_{n-1}, G/S'_n$$

except, again, for order, completing the proof.

7. Prove Th. 3.2.7.7.

PROOF:

G1: Since $a_1, a_2 \in A$ and $b_1, b_2 \in B$ imply $(a_1, b_1)(a_2, b_2) = (a_1a_2, b_1b_2)$ for which $a_1a_2 \in A$ and $b_1b_2 \in B$, it follows that $C = A \times B$ is closed.

G2: $A \times (B \times C) = (a_1, b_1) \times ((a_2, b_2) \times (a_3, b_3)) = (a_1(a_2a_3), b_1(b_2b_3))$
 $= ((a_1a_2)a_3, (b_1b_2)b_3) = ((a_1, b_1) \times (a_2, b_2)) \times (a_3, b_3) = (A \times B) \times C$

G3: $e \in A$ and $e' \in B$ implies $(e, e')(a, b) = (a, b)(e, e') = (a, b)$.

G4: $(a, b)^{-1}(a, b) = (a^{-1}, b^{-1})(a, b) = (e, e')$; $(a, b)(a, b)^{-1} = (a, b)(a^{-1}, b^{-1}) = (e, e')$.

8. If $A \times B = G$ in a group G , then A and B are normal in G , and $A \cap B = e$, $A \cup B = G$.

PROOF:

Let A and B be identified by two isomorphisms: $a \in A \leftrightarrow (a, 1) \in A \times B$ and $b \in B \leftrightarrow (1, b) \in A \times B$.

(i) Since $(a_1, b_2)^{-1}(a_2, 1)(a_1, b_1) = (a_1^{-1}, b_1^{-1})(a_2, 1)(a_1, b_1) = (a_1^{-1}a_2a_1, 1) \in A$, A is obviously a normal subgroup of $A \times B$; so is B likewise.

(ii) Since one and only one element which is simultaneously of the form $(a, 1)$ and $(1, b)$ is $(1, 1)$, it follows at once that $A \cap B = e$; otherwise, they are not even distinct, contradictory to what is implicit in the problem. Also, since $A \cup B$ contains every element of the form $(a, 1)(1, b) = (a, b)$, $A \cup B = A \times B = G$, completing the proof.

Note. Since the isomorphisms defined above also yield that $(a, 1)(1, b) = (a, b) = (1, b)(a, 1)$, it is evident that Df. 3.2.7.8a must be considered a theorem if the isomorphisms are first defined.

The converse of Prob. 8 is Prob. 10 below.

9. If G contains two subgroups A and B , either one of which is normal, then $A \cup B = AB$.

PROOF:

If A is normal, then $ba = bab^{-1}b = (bab^{-1})b = a'b$, and if B is normal, then $ba = aa^{-1}ba = a(a^{-1}ba) = ab'$. In either case the product can be written such that no b antecedes a , i.e. in general, $a_1a_2 \dots a_kb_{k+1} \dots b_n = ab$, where $a, a_i \in A$, $b, b_i \in B$. Hence every finite product of the form $g_1g_2 \dots g_n$ with $g_i \in A$ or B can always transform itself into the form ab , completing the proof.

10. Prove Th. 3.2.7.9.

PROOF:

Since A and B are both normal, they both contain an element $a^{-1}b^{-1}ab$ (since $a^{-1}(b^{-1}ab) \in A$ and $(a^{-1}b^{-1}a)b \in B$), which then must belong to $A \cap B = e$; hence $a^{-1}b^{-1}ab = e$, i.e. $ab = ba$.

Now, since $G = A \cup B = AB$, by Prob. 9 above, every element $g \in G$ can be written in the form $g = ab$, which is also unique, since $a_1b_1 = a_2b_2$ implies $a_2^{-1}a_1 = b_2b_1^{-1} \in A \cap B = e$, i.e. $a_1 = a_2$ and $b_1 = b_2$. Since $g_1 = a_1b_1$ and $g_2 = a_2b_2$ imply $g_1g_2 = a_1b_1a_2b_2 = (a_1a_2)(b_1b_2)$, and $g = ab$ implies $g \leftrightarrow ab$, the correspondence between G and $A \times B$ is isomorphic, completing the proof.

11. Prove Df. 3.2.7.8b by Th. 3.2.7.9.

PROOF:

If $n = 1$, then $G = G_1$, and the proof is trivial; so is it when $n = 2$, since it then becomes Prob. 8.

Assume that the proof has been completed for $n = k$. Then let $A' = \cup A_i$, $i = 1, 2, \dots, k$ and $B' = A_{k+1} = A_{k+1}$, which reduce the problem to setting up an isomorphism between G and its direct factors A_j , $j = 1, 2, \dots, k+1$, where (i) A' and B' are normal, (ii) $G = A' \cup B'$, and (iii) $A' \cap B' = e$, bringing down the problem to Prob. 8, by Th. 3.2.7.9, which completes the proof.

12. If $G = A \times B$, then $G/A \leftrightarrow B$ and $G/B \leftrightarrow A$, and if further $G' = A' \times B'$, where $A \leftrightarrow A'$ and $B \leftrightarrow B'$, then $G \leftrightarrow G'$.

PROOF:

(i) Since $g = ab$, for every $g \in G$, $a \in A$, $b \in B$, $H(g) = H(ab) = a$ is a homomorphism of G onto A and, by G1 of Prob. 7 above, $H(g_1g_2) = a_1a_2 = H(g_1)H(g_2)$. But $H(g) = e$ iff $g \in A$ since $H(g) = e$ iff $a = e$. Hence, by Th. 3.2.5.8, $G/B \leftrightarrow A$.

Likewise $H(g) = H(ab) = b$ is a homomorphism of G onto B , and $G/A \leftrightarrow B$.

(ii) Since two isomorphisms I and I' may be defined as $a \in A \leftrightarrow I(a) = a' \in A'$ and $b \in B \leftrightarrow I'(b) = b' \in B'$, which are given at the start, the correspondence $g = ab \leftrightarrow I(a)I'(b) = a'b' = g'$ is an isomorphism between G and G' .

13. Given nine permutations: $(1), (123), (132), (456), (465), (123)(456), (123)(465), (132)(456), (132)(465)$, prove that they form an Abelian group, then find the basis of the group.

PROOF:

Let $p = (123)$ and $q = (456)$; then $(1) = p^0q^0$, $(123) = p^1q^0$, $(132) = p^2q^0$, $(456) = p^0q^1$, $(465) = p^0q^2$, $(123)(456) = p^1q^1$, $(123)(465) = p^1q^2$, $(132)(456) = p^2q^1$, $(132)(465) = p^2q^2$, and in general $p^i q^j p^m q^n = p^{i+m} q^{j+n}$ and $p^i q^j p^{m-j} q^{n-j} = p^m q^n$, which provides G1-4. G5: Commutativity is also here, since p and q are mutually disjoint and, by §3.2.1, Prob. 16, $pq = qp$.

Since $p^i q^j = p^m q^n$ implies here $p^i = p^m$ and $p^j = q^n$, they are distinct as bases.

Note. This problem exemplifies a special case which, when generalized, yields the following theorem.

14. If A is an Abelian group, each element of which has as order a product of prime powers p_i , $i = 1, 2, \dots, n$, then A is of order $\prod p_i^{a_i}$ for some $a_i \geq 0$.

PROOF:

If each $a_i = 0$, then $\prod p_i^{a_i} = 1$ and $G = e$, in which case the theorem obviously holds.

If G is of order $n > 1$, then $G \neq e$ has an element g of order $q = \prod p_i^{b_i}$ for some $b_i \geq 0$, and the index of $\{g\}$ is $m = n/q < n$. Now, since each element of $A/\{g\}$ is of the form $h\{g\}$, where $h \in A$, and $h^r = e$ where $r = \prod p_i^{c_i}$ for some $c_i \geq 0$, $(h\{g\})^r = h^r\{g\} = \{g\}$ implies that r is divisible by the order of $h\{g\}$, which in turn implies that $A/\{g\}$ is of order $m = n/q$ such that each element has as order a product of prime powers p_i . But then, by the hypothesis of induction itself, $m = \prod p_i^{d_i}$ for some $d_i \geq 0$ and, in consequence, $n = qm = (\prod p_i^{b_i})(\prod p_i^{d_i}) = \prod p_i^{b_i + d_i}$.

Hence A is of order $\prod p_i^{a_i}$ where $a_i = b_i + d_i$ for some $b_i, d_i \geq 0$, $i = 1, 2, \dots, n$.

15. Prove Th. 3.2.7.10.

PROOF:

If $b_1, b_2 \in B$, then $b_1, b_2 \in A$ and, A being Abelian, $(b_1 b_2^{-1})^i = b_1^i (b_2^{-1})^i = e e^{-1} = e$, which implies $b_1 b_2^{-1} \in B$. Hence B is a subgroup of A and also normal, since A is Abelian. Likewise C is a normal subgroup of A .

Since i and j are relatively prime, there exist integers m and n such that $im + jn = 1$ (cf. Df. 4.1.2.3.15 and Th. 4.1.2.3.16) and, for every $a \in A$, $a = a^1 = a^{im+jn} = a^{im} a^{jn}$, which implies $a^{im} = c \in C$ (since $(a^{im})^j = a^{ijm} = a^{km} = (a^k)^m = e^m = e$) and $a^{jn} = b \in B$. Hence $a = bc \in BC$ and $A = BC = CB$, which satisfies the first condition of Df. 3.2.7.8a.

Let $g \in A$ such that $g \in B$ and $g \in C$; then $g^i = g^j = e$, which implies $g = g^{im+jn} = (g^i)^m (g^j)^n = e^m e^n = e$, which in turn implies that any element which is common in both B and C is the identity $e \in A$. Hence $B \cap C = e$, which satisfies the second condition of Df. 3.2.7.8a. Hence $A = B \times C$.

Furthermore, being relatively prime, i and j may be represented by $i = \prod p_u^{r_u}$, $u = 1, 2, \dots, x$, and $j = \prod q_v^{s_v}$, $v = 1, 2, \dots, y$, where $p_u \neq q_v$ are primes (cf. Th. 4.1.2.3.17), while, by Prob. 14 above, B is of order $\prod p_u^{t_u}$ and C of order $\prod q_v^{w_v}$. Now, since the third condition of Df. 3.2.7.8a demands that $a = bc$ be unique, and consequently that the order k of A be the product of the order of B and C , it must be the case that $k = (\prod p_u^{t_u})(\prod q_v^{w_v}) = (\prod p_u^{r_u})(\prod q_v^{s_v}) = ij$. But then, since the factorization of an integer into prime power factors is unique (cf. Th. 4.1.2.3.17), it follows that $t_u = r_u$ and $w_v = s_v$. Hence B is uniquely of order $i = \prod p_u^{r_u}$ and C of order $j = \prod q_v^{s_v}$ where $k = ij$, completing the proof.

Supplementary Problems

Part 3

- 3.1. How many symmetries are possessed by the regular tetrahedron, the regular hexahedron (i.e. cube), the regular octahedron, the regular dodecahedron, and the regular icosahedron?
- 3.2. Find the number of all rotations, including the original position, with respect to the regular polyhedrons of Prob. 3.1 above.
- 3.3. If a set C satisfies the following four axioms, then C is a commutative group:
- | | |
|---|----------------|
| C1 \equiv G1 (cf. Df. 3.1.1.1) | C3 \equiv G3 |
| C2. $a(bc) = (ba)c$ for every $a, b, c \in C$ | C4 \equiv G4 |

- 3.4. The following six functions form a group under the operative rule: $f_i f_j(x) = f_i(f_j(x))$, $i, j = 1, 2, \dots, 6$, (cf. Df. 2.2.2.14):

$$\begin{array}{lll} f_1(x) = x & f_3(x) = 1 - x & f_5(x) = (x - 1)/x \\ f_2(x) = 1/x & f_4(x) = 1/(1 - x) & f_6(x) = x/(x - 1) \end{array}$$

- 3.5. If G is a group and $a, b, c_1, c_2, \dots, c_n \in G$, then,

- (i) $aabb = ab$ implies $b = a^{-1}$;
- (ii) $(bab^{-1})^n = ba^n b^{-1}$.
- (iii) $(ac_1 a^{-1})(ac_2 a^{-1}) \dots (ac_n a^{-1}) = a(c_1 c_2 \dots c_n) a^{-1}$.

- 3.6. The inverse of the inverse permutation P^{-1} of the original permutation P is P .

- 3.7. A polynomial $P(x_1, x_2, \dots, x_n)$ in n indeterminates x_i is called *symmetric* (cf. Df. 5.2.2.1) if it is invariant under the symmetric group of all permutations of its subscripts. Prove the following polynomials are symmetric, considering $x_1 = a$, $x_2 = b$, $x_3 = c$:

- (i) $(a + b - c)(b + c - a)(c + a - b)$
- (ii) $(a - b)^2(a - c)^2 + (a - b)^2(b - c)^2 + (a - c)^2(b - c)^2$
- (iii) $(a - b)^2(a + b - c) + (b - c)^2(b + c - a) + (c - a)^2(c + b - a)$.

- 3.8. The symmetric polynomial of Prob. 3.7 above is called *alternating* if it remains the same except for its signs under all permutations of its subscripts. Prove that the following polynomials are alternating:

- (i) $x_1(x_2 - x_3) + x_2(x_3 - x_1) + x_3(x_1 - x_2)$
- (ii) $(x_1 - x_2)(x_1 - x_3)(x_1 - x_4)(x_2 - x_3)(x_2 - x_4)(x_3 - x_4)$.

- 3.9. Prove that the product of a symmetric polynomial and an alternating polynomial is an alternating polynomial, and that the product of two alternating polynomials is a symmetric polynomial.

- 3.10. If G is a permutation group of n numbers: $1, 2, \dots, n$, where $n \geq 3$, to which $(n - 2)$ cyclic permutations: $(123), (124), \dots, (12n)$, belong, then G is either a symmetric group or an alternating group.

- 3.11. Verify that any symmetric group of degree greater than three is not commutative.

- 3.12. If I_1 is a subgroup of the additive group I of all integers, and if I_1 does not contain 0, then the elements of I_1 are the multiples of the least positive integer of I_1 .

- 3.13. The set S of all the common multiples of n integers a_1, a_2, \dots, a_n forms an additive group whose elements are the multiples of the least common multiple of the given integers.

- 3.14. If d is the greatest common divisor of n integers a_1, a_2, \dots, a_n , then there exists a set of integers b_1, b_2, \dots, b_n such that $d = a_1 b_1 + a_2 b_2 + \dots + a_n b_n$.

- 3.15. Given an infinite cyclic group G whose generator is g , let G_1 be a subgroup of G generated by g^m , and G_2 a subgroup of G generated by g^n ; then, there exists a subgroup G_3 of G generated by g^d , where d is the greatest common divisor of m and n , such that the elements of G_3 are of the form $a_i b_j$, where $a_i \in G_1$, $b_j \in G_2$.

- 3.16. Find all the proper subgroups of the symmetric group $S_3 = \{a, b, c\}$.

- 3.17. Given a polynomial $ab + cd$, verify that there exists a set M of eight permutations $(1), (ab), (cd), (ab)(cd), (ac)(bd), (ad)(bc), (acbd), (adbc)$ such that the polynomial is invariant under the permutations, and that the set of the permutations is a subgroup of $S_4 = \{a, b, c, d\}$.

- 3.18. Verify that the polynomial $ab + cd$ of Prob. 3.17 above is changed to $ad + bc$ by a transposition (bc) , and that $ad + bc$ is invariant under the eight permutations of $M(bc)$. Likewise $ac + bd$ is invariant under the eight permutations of $M(bd)$.

- 3.19. Prove that, in the context of Prob. 3.17-18 above,

$$S_4 = M \cup M(bc) \cup M(bd)$$

where $M \leftrightarrow (ab + cd)$, $M(bc) \leftrightarrow (ad + bc)$, and $M(bd) \leftrightarrow (ac + bd)$.

- 3.20. Reconsider Prob. 3.17-18 in terms of conjugacy.
- 3.21. (a) Find the order of the octahedral group by a right decomposition (cf. Df. 3.2.2.6).
 (b) Reinterpret Prob. 3.2 above in terms of right (or left) decompositions.
- 3.22. Interpret the rotation group of the square, first in terms of cosets, then in terms of conjugate classes.
- 3.23. Construct a subgroup of the multiplicative group \bar{R} of all real numbers, into which a homomorphism of R may take place.
- 3.24. Verify the homomorphism of the rotation group of a regular $2n$ -gon onto the symmetric group of degree n .
- 3.25. If n is the integral divisor of an integer m , then the cyclic group of order m is homomorphic onto the cyclic group of order n .
- 3.26. If $g_1G_1, g_2G_1, \dots, g_nG_1$ exhaust the left-cosets of a subgroup G_1 of a group G , then the following correspondence, for any $a \in G$,

$$a \rightarrow \begin{pmatrix} g_1G_1 & g_2G_1 & \dots & g_nG_1 \\ ag_1G_1 & ag_2G_1 & \dots & ag_nG_1 \end{pmatrix}$$

yields a homomorphism of G onto a permutation group of order n .

- 3.27. Verify that the tetrahedral group is isomorphic to the alternating group of degree 4.
- 3.28. Given an isomorphism between the infinite cyclic group C whose generator is c and the additive group A of all integers, where the pattern of the isomorphism is prescribed by $c^a \in C \leftrightarrow a \in A$, find a subgroup C_2 of C which is isomorphic to a subgroup C_1 of C which consists of all the multiples of a positive integer n .
- 3.29. The normal subgroup of a group G is always a join of some conjugate classes of G .
- 3.30. Given the symmetric group S_3 , find its subgroups, cosets, and normal subgroups.
- 3.31. Prove that the symmetric group S_3 has six, and only six, inner automorphisms.
- 3.32. If A_1, A_2, A_3 are the three symmetric axes which connect the three pairs of the opposite vertices of the regular octahedron, then the octahedral group is homomorphic onto the symmetric group of A_1, A_2, A_3 and its kernel is the four group V_4 .
- 3.33. If a group G of order $2p$, where p is a prime, has a normal subgroup N of order 2, then G is a commutative group.
- 3.34. If a group G of order 6 is not commutative, it is isomorphic to the symmetric group S_3 .
- 3.35. Given two groups A and B , their direct product $A \times B$ contains two subgroups, one isomorphic to A and the other isomorphic to B .
- *3.36. If A and B are two groups, $A \times B$ is their direct product, and e_a and e_b are their identity subgroups respectively, then:
- (i) $A \times e_b = C$ and $e_a \times B = D$ have one, and only one, element in common, viz. the identity.
 - (ii) Every element of C commutes with every element of D .
 - (iii) $cd = p \in A \times B$ is unique for every $c \in C$ and $d \in D$.
- *3.37. If $A \times B \leftrightarrow C$, where A and B are subgroups of a group C , then:
- (i) A and B have one, and only one, element in common, viz. the identity.
 - (ii) Every element of A commutes with every element of B .
 - (iii) $ab = c \in C$ is unique for every $a \in A$ and $b \in B$.
- *3.38. If A, B, C , and D are subgroups of a group G , and if C and D are normal subgroups of A and B , then

$$((A \cap B)C)/((A \cap D)C) \leftrightarrow ((B \cap A)D)/((B \cap C)D)$$

Rings

§4.1.1 Rings in General

Df. 4.1.1.1 A module (i.e. an additive Abelian group) R is a *ring* if R is also a semi-group under multiplication and, further, satisfies right and left distributions under addition.

Stated in detail, R satisfies the following eight axioms, for any elements $a, b, c \in R$:

- | | |
|--|--|
| R1: Additive closure. | $a, b \in R$ implies $a + b \in R$. |
| R2: Additive associativity. | $a + (b + c) = (a + b) + c$ |
| R3: Additive identity. | $a + 0 = 0 + a = a$ |
| R4: Additive inverse. | $a + (-a) = (-a) + a = 0$ |
| R5: Additive commutativity. | $a + b = b + a$ |
| R6: Multiplicative closure. | $a, b \in R$ implies $ab \in R$. |
| R7: Multiplicative associativity. | $a(bc) = (ab)c$ |
| R8: Additive distributivity: | $a(b + c) = ab + ac, (b + c)a = ba + ca$ |

The additive identity 0, called the *zero* of the ring, is unique; so is the additive inverse of a , denoted by $-a$ and called the *negative* (or more simply, *minus*) a . (Cf. Prob. 1.) Subtraction is thus possible and unique in R , since the equation $a + x = b$ has a unique solution in any additive Abelian group (cf. Prob. 13 below).

These two binary operations of addition (including subtraction) and multiplication resemble the familiar rational operations of elementary algebra, however, only to a certain extent; for the absence of **G3-4** under multiplication results in the conspicuous presence of possible divisors of zero.

Df. 4.1.1.2 R is a ring with *divisors of zero* (or *zero-divisors*) if $x \cdot y = 0$ when $x, y \in R$, $x \neq 0$, $y \neq 0$, x being a *left* and y a *right zero-divisor*. Otherwise, i.e. if $x \cdot y = 0$ implies $x = 0$ or $y = 0$, the ring is called a ring without divisors of zero. Zero itself is considered a divisor of zero, if only for the sake of expediency.

Also, as is quite explicit in Df. 4.1.1.1, R may not commute under multiplication, may not have a unit element, and may not have inverses for its elements even if it has a unit element. But then, of course, R may at times satisfy **G3-5** under multiplication, and if it does, it needs the following additional definitions.

Df. 4.1.1.3 R is a ring with *unity* if it satisfies **G3** under multiplication, having a multiplicative identity e , called the unity of R .

Th. 4.1.1.4 If an element x of a ring R with unity has a multiplicative inverse, denoted by x^{-1} , then the inverse is unique. (Cf. Prob. 1.)

Df. 4.1.1.5 R is a *commutative ring*, if it satisfies **G5** under multiplication, i.e. $xy = yx$ for every $x, y \in R$; otherwise, i.e. $xy \neq yx$ for at least two elements $x, y \in R$, R is called a *noncommutative ring*.

Df. 4.1.1.6 A complex C of a ring R is a *subring* if C is also a ring, satisfying the axioms **R1-8**.

A commutative ring may or may not have a unity; e.g. the set I of all integers is a commutative ring with unity, while its subring I_e of all even integers is a commutative ring without unity.

A ring R is then, first and foremost, a module, i.e. an Abelian group under addition; hence many theorems proved for Abelian groups hold with slight modifications for R , as is best exemplified in Th. 4.1.1.7.

Th. 4.1.1.7 A complex C of a ring R is a subring iff $a-b \in C$ and $ab=ba \in C$ for every $a, b \in C$. (Cf. Th. 3.2.1.2 and also Prob. 2 below.)

Just as some alternative definitions were available for subgroups, subrings may be proved in some other ways; for example:

Th. 4.1.1.8 If a complex C of a ring R satisfies **R1, 6**, and if $c \in C$ implies $-c \in C$, then C is a subring of R . (Cf. Prob. 3.)

The relations among rings or among rings and subrings reintroduce here the familiar concepts of homomorphisms and isomorphisms (cf. §2.2.2 and §3.1.3).

Df. 4.1.1.9 If the mapping of a ring R onto or into a ring R' : $a \in R \rightarrow a' \in R'$ and $b \in R \rightarrow b' \in R'$ implies $a+b \in R \rightarrow a'+b' \in R'$ and $ab \in R \rightarrow a'b' \in R'$, then the mapping is a *homomorphism*, while it is an *isomorphism* if the mapping is one-one, i.e. $a \in R \leftrightarrow a' \in R'$ and $b \in R \leftrightarrow b' \in R'$ imply $a+b \in R \leftrightarrow a'+b' \in R'$ and $ab \in R \leftrightarrow a'b' \in R'$.

Again, many homomorphisms or isomorphisms established for Abelian groups may be modified for rings.

***Th. 4.1.1.10** If S is a homomorph of a ring R in R' when R' itself is a homomorph of R , then S is a subring of R' . (Cf. Prob. 5.)

***Th. 4.1.1.11** Given two rings R_1 and R_2 with no elements in common, where R_1 contains a subring S_1 isomorphic to R_2 , there exists a ring R_3 which is isomorphic to R_1 and contains R_2 as a subring. (Cf. Prob. 7.)

This fundamental theorem of rings, sometimes called a general replacement theorem, asserts the existence of a ring which is isomorphic to the one constructed to have prescribed properties and which actually contains the given rings (cf. Th. 4.2.1.2-7).

Rings in general have the following properties.

Th. 4.1.1.12 (Cancellation under addition). For every $a, b, c \in R$, $a+c = b+c$ or $c+a = c+b$ implies $a = b$. (Cf. Prob. 12.)

Th. 4.1.1.13 For every $a \in R$, $a \cdot 0 = 0 \cdot a = 0$. (Cf. Prob. 14.)

Th. 4.1.1.14 For every $a, b \in R$,

- | | |
|-----------------------|---|
| (i) $-(-a) = a$ | (iv) $a(-b) = (-a)b = -(ab) \equiv -ab$ |
| (ii) $-(a+b) = -a-b$ | (v) $(-a)(-b) = ab$. (Cf. Prob. 16.) |
| (iii) $-(a-b) = -a+b$ | |

Th. 4.1.1.15 For every $a, b \in R$ and every $m, n \in I^+$ (positive integers),

$$(i) \quad a^m \cdot a^n = a^{m+n} \quad (ii) \quad (a^m)^n = a^{mn}$$

and if R is commutative,

$$(iii) \quad (ab)^n = a^n b^n. \quad (\text{Cf. Prob. 20.})$$

Th. 4.1.1.16 For every $a, b \in R$ and every $m, n \in I$ (any integers),

$$(i) \quad n(a+b) = na + nb \quad (iii) \quad n(ab) = (na)b = a(nb). \quad (\text{Cf. Prob. 21.})$$

$$(ii) \quad (m+n)a = ma + na$$

Solved Problems

1. The zero of a ring R is unique; so are the additive inverse $-a$ and, in case R is with unity, the multiplicative inverse a^{-1} of $a \in R$.

PROOF:

- (i) Let $0, 0' \in R$ such that $a+0 = 0+a = 0$ and $a+0' = 0'+a = 0'$. Then, by **R3**, $0'+0 = 0+0' = 0$ and $0+0' = 0'+0 = 0'$. But, by **R5**, $0'+0 = 0+0'$. Hence $0 = 0'$, proving that 0 must be unique.
- (ii) Suppose $b \in R$ such that $b \neq 0$ and yet $a+b = 0$ for every $a \in R$. Then, since $-a \in R$ and $a+(-a) = 0$, it follows that, by **R1**, $(-a)+(a+b) = (-a)+(a+(-a))$ and, by **R2**, $((-a)+a)+b = ((-a)+a)+(-b)$. Hence, by **R4**, $0+b = 0+(-a)$, i.e. $b = -a$, proving that the additive inverse is unique.

The uniqueness of a^{-1} can be proved likewise, completing the proof.

Note the similarity between this problem and §3.1.1, Prob. 1; the only difference, in fact, is the operators, viz. “ \ast ” or “ \circ ” there and “ $+$ ” here.

2. Prove Th. 4.1.1.7.

PROOF:

- (i) If the complex C is a subring of R , then $a, b \in C$ implies $ab \in C$, by **R6**, and also $-b \in C$, by **R4**; hence, by **R1**, $a-b \in C$, proving that $ab \in C$ and $a-b \in C$ are necessary conditions for C to be a subring of R .
- (ii) Conversely, these conditions are sufficient for C to be a subring. For $ab=ba \in C$ already satisfies (more than) **R6**, and $a \in C$ and $a-b \in C$ imply $a-a = 0 \in C$, providing **R4**, which in turn yields $0+(-a) = -a \in C$, setting up **R3**. Furthermore, $-a \in C$ for a implies $-b \in C$ for b and consequently $a-(-b) = a+b \in C$, establishing **R1**. **R1** and **R6** imply **R2** and **R7** respectively, as is trivial by now. Finally, **R1** and $ab = ba \in C$ uniquely determine $c(a+b)$ and $(a+b)c$ for every $c \in C$, preparing for **R8**, which completes the proof.

Note that $c(a+b) = ca+cb$ and $(a+b)c = ac+bc$ are a matter, not of deduction, but of definition; it is quite legitimate and consistent, however, to define them as such at this juncture, since $c(a+b)$ and $(a+b)c$ are uniquely determined.

3. Prove Th. 4.1.1.8.

PROOF:

From **R1** and $-c \in C$ for $c \in C$ it follows at once that $c+(-c) = 0 = (-c)+c$, providing **R4**, and the rest can be carried out as above.

Note, again, that the introduction of “0” is a matter, not entirely of deduction, but more of definition; all that can be purely deduced from “**R1** and $-c \in C$ for $c \in C$ ” is that “ $c+(-c)$ ”, whatever it may be, is unique, for which “0” may always consistently stand.

4. Prove that the complex C of the additive identity, i.e. zero, alone is a subring of a ring R .

PROOF:

Since $0 + 0 = 0$ and $0 \cdot 0 = 0$, it satisfies **R1-2**, **R6-7**, and also **R3-4**, since it is the identity and the inverse of itself. Finally, $0(0 + 0) = (0 + 0)0 = 0$, satisfying **R8**.

5. Prove Th. 4.1.1.10.

PROOF:

Let H be the homomorphism of mapping of R onto R' ; then, since $R' = H(R)$ is a homomorph of the additive Abelian group R , S in R' is clearly a subgroup of the additive Abelian group R' , satisfying **R1-5**. Moreover, if $a, b \in R$ are the preimages in R of any two elements $a', b' \in S$, corresponding to a and b respectively, then $H(ab) = H(a)H(b) = a'b' \in S$, satisfying **R6** for S , and consequently also **R7** for S . Likewise, if $a, b, c \in R$ are the preimages of any three elements $a', b', c' \in S$, corresponding to a, b, c respectively, then $H(a(b+c)) = H(a)H(b+c) = a'(b'+c') \in S$, satisfying **R8**. (Or still further, $H(a(b+c)) = H(ab+bc) \rightarrow a'(b'+c') = (a'b' + a'c') \in S$.)

This completes the proof.

6. If S is a set of elements in a 1-1 correspondence with the elements of a ring R , then S can afford two binary operations, i.e. addition and multiplication, such that S becomes a ring isomorphic to R .

PROOF:

The 1-1 correspondence between R and S , e.g. $a \in R \leftrightarrow a' \in S$ and $b \in R \leftrightarrow b' \in S$, becomes an isomorphism I if it is defined, as is but logical, that $I(a) = a'$, $I(b) = b'$. Then

$$a+b \in R \leftrightarrow I(a+b) = I(a) + I(b) = a' + b' \in S$$

and

$$a \cdot b \in R \leftrightarrow I(a \cdot b) = I(a) \cdot I(b) = a' \cdot b' \in S$$

Hence S will, and can, satisfy all of **R1-8**, proving that S is a ring isomorphic to R .

7. Prove Th. 4.1.1.11.

PROOF:

When schematized, the problem is to prove the following relation:

$$\begin{array}{ccc} R_1 & \supset & S_1 \\ \updownarrow & & \updownarrow \\ R_3 & \supset & R_2 \end{array}$$

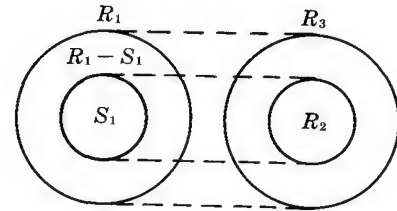


Fig. 3.1.1a

whose diagram is given at right.

Let $R_3 = (R_1 - S_1) \cup R_2$, and define a mapping M such that $a \in (R_1 - S_1)$ implies $M(a) = a$ and also $s \in S_1$ implies $M(a) = I(a)$, where I is by hypothesis the isomorphic mapping of S_1 onto and into R_2 . Moreover, since M is then the 1-1 mapping of $R_1 - S_1$ into itself and, R_1 and R_2 (or more narrowly, $R_1 - S_1$ and R_2) having no elements in common, M is in fact an isomorphism between R_1 and R_3 .

Furthermore, the binary operations of R_3 are defined by M , viz.,

$$\begin{array}{lcl} a+b \in R_1 & \leftrightarrow & M(a+b) = M(a) + M(b) = a' + b' \in R_3 \\ \text{and} & & \\ a \cdot b \in R_1 & \leftrightarrow & M(a \cdot b) = M(a) \cdot M(b) = a' \cdot b' \in R_3, \end{array} \quad (1)$$

which proves R_3 to be a ring (cf. Prob. 6 above).

Finally, since I is the isomorphic mapping of S_1 onto R_2 ,

$$\begin{array}{lcl} a+b \in S_1 & \leftrightarrow & I(a+b) = I(a) + I(b) = a' + b' \in R_2 \\ \text{and} & & \\ a \cdot b \in S_1 & \leftrightarrow & I(a \cdot b) = I(a) \cdot I(b) = a' \cdot b' \in R_2 \end{array} \quad (2)$$

where (1) actually coincides with (2), since $M(a), M(b) \in R_2$ in (1) implies $a, b, a+b, ab \in S_1$. Hence R_2 is also a ring and, as a matter of fact, a subring of R_3 according to the initial assumption, completing the proof.

8. Prove that residue classes modulo 2, 3, 4 are rings.

PROOF:

It is self-explanatory through the following tables:

(i)

+	0	1
0	0	1
1	1	0

•	0	1
0	0	0
1	0	1

(ii)

+	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

•	0	1	2
0	0	0	0
1	0	1	2
2	0	2	1

(iii)

+	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

•	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	0	2
3	0	3	2	1

Observe that the last ring, for instance, is a ring with unity, but not every nonzero element (e.g. 2) has an inverse; since $2 \cdot 2 = 0$, the ring has divisors of zero. This ring is also a commutative ring, as can be readily verified by the table.

9. Generalize Prob. 8; viz. a residue class modulo m , where m is any positive integer, is a ring.**PROOF:**

The residue class C modulo m is an additive Abelian group (cf. §3.2.6, Prob. 11), satisfying **R1-5**. Furthermore, by Th. 3.2.6.8, C satisfies **R6** and consequently **R7**, proving itself to be a semi-group under multiplication, from which **R8** also follows, viz. for every $a, b, c \in C$, $a(b+c) \equiv ab+ac \pmod{m}$ and $(a+b)c \equiv ac+bc \pmod{m}$, (cf. Th. 3.2.6.8-9).

Hence C is a ring, and as has already been shown above, generally a commutative ring with unity and possibly with divisors of zero.

10. Examine whether the following sets form rings: (i) the set of all natural numbers; (ii) the set of all integers; (iii) the set of all positive rational numbers; (iv) the set of all rational numbers; (v) the set of all real numbers of the form $x + y\sqrt{2}$, where x and y are integers; (vi) the set of all real numbers of the form $x + y\sqrt[3]{2} + z\sqrt[3]{4}$, where x, y, z are rational numbers; (vii) the set of all real numbers; (viii) the set of all complex numbers of the form $x + yi$, where x and y are integers; (ix) the set of all complex numbers; (x) the set of all real-valued continuous functions on the interval $-1 \leq x \leq 1$, where $(f+g)(x) = f(x) + g(x)$ and $(fg)(x) = f(x)g(x)$.**PROOF:**

The sets of (i) and (iii) are obviously not rings; for, in the first place, they are not even additive Abelian groups, unable to satisfy **R4**, for instance.

All others form rings in various ways, satisfying at least **R1-8** or more, e.g. with unity or without zero divisors or with commutativity.

The sets of (ii), (iv)-(ix) are all commutative rings with unity and without divisors of zero, as can be verified without difficulty (cf. e.g. §3.1.1, Prob. 10). Furthermore, each of them is a subring of (ix); also, in detail, each of (v)-(viii) is a proper subring of (ix), each of (iv)-(vi) is a proper subring of (vii), and (ii) is a proper subring of (iv).

The set of (x) forms also a commutative ring and, if either $f(x) = 1$ or $g(x) = 1$, it is a ring with unity, and possibly with divisors of zero; e.g. if $f(x) = \max(0, x)$ and $g(x) = \max(0, -x)$, then both functions are distinct from zero, yet their product is zero.

11. The set C of the integral couples (x, y) forms a ring if they are operated on as follows:

$$(x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2), \quad (x_1, y_1)(x_2, y_2) = (x_1x_2, y_1y_2);$$

so does the set T of integral triples (x, y, z) if they comply with the following operative rules:

$$(x_1, y_1, z_1) + (x_2, y_2, z_2) = (x_1 + x_2, y_1 + y_2, z_1 + z_2), \quad (x_1, y_1, z_1)(x_2, y_2, z_2) = (x_1x_2, x_2y_1 + y_2z_1, z_1z_2)$$

PROOF:

- (i) Since **R1** and **R6** for the set C are already given by the operative rules, **R2** and **R7** follow immediately; viz.,

$$\begin{aligned} (x_1, y_1) + ((x_2, y_2) + (x_3, y_3)) &= (x_1 + (x_2 + x_3), y_1 + (y_2 + y_3)) = ((x_1 + x_2) + x_3, (y_1 + y_2) + y_3) \\ &= ((x_1, y_1) + (x_2, y_2)) + (x_3, y_3) \end{aligned}$$

$$\text{and likewise } (x_1, y_1)((x_2, y_2)(x_3, y_3)) = ((x_1, y_1)(x_2, y_2))(x_3, y_3).$$

R3 is obviously satisfied by $(0, 0)$, and **R4** by $(x, y)^{-1} = (-x, -y)$.

As for **R8**, it is also satisfied, since

$$\begin{aligned} (x_1, y_1)((x_2, y_2) + (x_3, y_3)) &= (x_1, y_1)(x_2 + x_3, y_2 + y_3) = (x_1x_2 + x_1x_3, y_1y_2 + y_1y_3) \\ &= (x_1, y_1)(x_2, y_2) + (x_1, y_1)(x_3, y_3) \end{aligned}$$

$$\text{Similarly } ((x_1, y_1) + (x_2, y_2))(x_3, y_3) = (x_1, y_1)(x_3, y_3) + (x_2, y_2)(x_3, y_3).$$

The set C is a ring with unity $(1, 1)$ without zero divisors, and also commutative, as can be readily verified.

- (ii) The set T satisfies also all of **R1-8**; e.g., for **R8**:

$$\begin{aligned} (x_1, y_1, z_1)((x_2, y_2, z_2) + (x_3, y_3, z_3)) &= (x_1, y_1, z_1)(x_2 + x_3, y_2 + y_3, z_2 + z_3) \\ &= (x_1(x_2 + x_3), (x_2 + x_3)y_1 + (y_2 + y_3)z_1, z_1(z_2 + z_3)) \\ &= (x_1x_2, y_2y_1 + y_2z_1, z_1z_2) + (x_1x_3, x_3y_1 + y_3z_1, z_1z_3) \\ &= (x_1, y_1, z_1)(x_2, y_2, z_2) + (x_1, y_1, z_1)(x_3, y_3, z_3) \end{aligned}$$

The set T is a ring with unity $(1, 0, 1)$, with divisors of zero (e.g. $(0, 1, 0)(1, 0, 1) = (0, 0, 0)$), and noncommutative (e.g. $(0, 0, 1)(0, 1, 0) = (0, 1, 0) \neq (0, 0, 0) = (0, 1, 0)(0, 0, 1)$).

Note that, since x, y, z are all integers, there exists multiplicative inverses iff $x = \pm 1$ and $z = \pm 1$ (since $(x_1, y_1, z_1)(x_2, y_2, z_2) = (1, 0, 1)$ implies $x_1x_2 = 1$ and $z_1z_2 = 1$).

12. For every $a, b, c \in R$, $a + c = b + c$ implies $a = b$; likewise, $c + a = c + b$ implies $a = b$.

PROOF:

Since, by **R4**, there must be an element $c^{-1} \in R$ such that $c + c^{-1} = 0$, it follows that if $a + c = b + c$, then, by **R1-3**, $(a + c) + c^{-1} = (b + c) + c^{-1} \rightarrow a + (c + c^{-1}) = b + (c + c^{-1}) \rightarrow a + 0 = b + 0 \rightarrow a = b$.

Likewise, $c + a = c + b$ implies $a = b$.

Note. This is the Cancellation Law for a ring R , which indeed must exist, since R is after all an additive Abelian group, for which the cancellation law does exist (cf. Th. 2.1.1.3).

Given this law first, the proof of Prob. 1 can be considerably simplified; e.g. if both x and y are the additive inverse of $a \in R$, then $a + x = 0$ and $a + y = 0$, which by this law implies $x = y$, proving the uniqueness of the inverse.

13. If $a, b \in R$, then $a + x = b$ has in R a solution $x = b - a$, which is unique.

PROOF:

Since $a + (b - a) = a + (-a + b) = (a - a) + b = b$, it is obviously the case that $x = b - a$. Furthermore, this is the only solution for $a + x = b$; for, if also $a + y = b$, then $a + x = a + y$ and, by the cancellation law, $x = y$, completing the proof.

14. Prove Th. 3.1.1.13.

PROOF:

Since $a + 0 = a$ by **R4**, it follows that $a(a + 0) = a \cdot a$ by **R6**, while also, by **R8**, $a(a + 0) = a \cdot a + a \cdot 0$. Hence $a \cdot a + a \cdot 0 = a \cdot a$ and, by Prob. 12 above, $a \cdot 0 = a \cdot a - a \cdot a = 0$.

Likewise $0 \cdot a = 0$, completing the proof.

15. Prove that $(a+b)^{-1} = -(a+b)$, and that $(ab)^{-1} = b^{-1}a^{-1}$ if $a^{-1} \in R$ and $b^{-1} \in R$ under multiplication.

PROOF:

- (i) $(a+b) + ((-a)+(-b)) = ((a+b)+(-a)) + (-b) = ((a+(-a))+b) + (-b) = (0+b) + (-b) = b + (-b) = 0$. Hence $(-a)+(-b)$ is the additive inverse of $a+b$; but, the inverse being unique by **R4**, $(-a)+(-b) = -(a+b)$, i.e. the additive inverse of $a+b$ is $-(a+b)$ which may notationally be written as $(-a)+(-b)$ (or, as will be proved below, $-a-b = -(a+b)$).
- (ii) $(ab)(b^{-1}a^{-1}) = a(bb^{-1})a^{-1} = aea^{-1} = aa^{-1} = e$, proving that the inverse of ab , i.e. $(ab)^{-1}$, is uniquely $b^{-1}a^{-1}$; the uniqueness is provided by Prob. 1, (ii).

16. Prove Th. 4.1.1.14.

PROOF:

- (i) $(-a) + (-(-a)) = 0$ and $(-a) + a = 0$, by **R4**; hence $(-a) + (-(-a)) = (-a) + a$ and, by cancellation, $-(-a) = a$.

Second proof. $(0+a) + (-a) = 0 + (a+(-a)) = 0 + 0 = 0$; hence, by Prob. 13, $(0-a) = 0 - (-a)$, i.e. $a = -(-a)$.

- (ii) Let $a+b = c \in R$, by **R1**; then, by **R4**, $c + (-c) = 0$ and, by substitution, $(a+b) + (-a-b) = 0$, while $(a+b) + (-a-b) = (a+b) + (-b-a) = a + (b+(-b)) + (-a) = a + 0 + (-a) = a + (-a) = 0$. Hence $(a+b) + (-a-b) = (a+b) + (-a-b)$ and, by cancellation, $-(a+b) = -a-b$.
- (iii) Replace b by $-b$ in (ii), and $-(a+(-b)) = -a-(-b)$; then, by (i), $-(a-b) = -a+b$.
- (iv) $((-a)+a)b = (-a)b + ab$, by **R8**, and also $((-a)+a)b = 0 \cdot b = 0$, by **R4** and Th. 4.1.1.13. Hence $(-a)b + ab = 0$ and, by Prob. 13, $(-a)b = -(ab)$. Likewise $a(b+(-b)) = ab + a(-b) = 0$ and $a(-b) = -(ab)$. Hence $(-a)b = a(-b) = -(ab)$, and notationally, by definition, $-(ab) = -ab$.
- (v) $(ab+a(-b)) + (-a)(-b) = a(b+(-b)) + (-a)(-b) = a \cdot 0 + (-a)(-b) = (-a)(-b)$, by **R8, R4**, while $ab + (a(-b) + (-a)(-b)) = ab + (a+(-a))(-b) = ab + 0 \cdot (-b) = ab$, again by **R8, R4**. But $(ab+a(-b)) + (-a)(-b) = ab + (a(-b) + (-a)(-b))$, by **R2**. Hence $(-a)(-b) = ab$.

Second proof. By (iv), $(-a)(-b) = -(a(-b)) = -(-(ab))$, which, by (i), equals ab . Hence $(-a)(-b) = ab$.

17. Prove (i) $(a-b)-c = a-(b+c)$, (ii) $a(b-c) = ab-ac$.

PROOF:

- (i) Since $-(a+b) = -a-b$ (cf. Prob. 16, (ii) above), $(a-(b+c)) + c = a-b-c+c = a-b+0 = a-b$. Hence $a-(b+c) = (a-b)-c$, by Prob. 12.
- (ii) $a(b+(-c)) = ab + a(-c) = ab-ac$, by **R8** and Prob. 16, (iv).

18. Prove the following properties of differences: for every $a, b, c, d \in R$,

- (i) $a-b = c-d$ iff $a+d = b+c$ (iii) $(a-b)-(c-d) = (a+d)-(b+c)$
(ii) $(a-b)+(c-d) = (a+c)-(b+d)$ (iv) $(a-b)(c-d) = (ac+bd)-(ad+bc)$

PROOF:

- (i) If $a-b = c-d$, then adding $(b+d)$ to both sides of the equation, $(a-b) + (b+d) = (c-d) + (b+d)$ which, when simplified by **R2, 4, 5**, becomes $a+d = b+c$. Conversely, if $a+d = b+c$, then adding $((-b)+(-d))$ to both sides of the equation, $(a+d) + ((-b)+(-d)) = (b+c) + ((-b)+(-d))$, and simplifying likewise, $a-b = c-d$.
- (ii) $((a-b)+(c-d)) + (b+d) = (a-b) + ((c-d)+(b+d)) = (a-b) + (c+b) = a+c$. Hence, by Prob. 12, $(a-b) + (c-d) = (a+c) - (b+d)$.
- (iii) Likewise $(a-b) - (c-d) = (a+d) - (b+c)$.
- (iv) By left-distribution, then right-distribution, and **R2, 5**, $(a-b)(c-d) = (a-b)c - (a-b)d = (ac-bc) - (ad-bd) = ac-bc-ad+bd = (ac+bd) - (ad+bc)$. Or what is the same, first by right-distribution, then left-distribution and **R2, 5**, $(a-b)(c-d) = a(c-d) - b(c-d) = ac-ad-bc+bd = (ac+bd) - (ad+bc)$.

19. Generalize **R2**, **R7**, and **R8**.**PROOF:**

- (i) Since, by
- R2**
- ,
- $a_1 + (a_2 + a_3) = (a_1 + a_2) + a_3 = a_1 + a_2 + a_3 = \sum_{i=1}^3 a_i$
- , then

$$\begin{aligned}
 \left(\sum_{i=1}^r a_i \right) + \left(\sum_{j=r+1}^s a_j \right) &= (a_1 + a_2 + \cdots + a_r) + (a_{r+1} + a_{r+2} + \cdots + a_s) \\
 &= (a_1 + a_2 + \cdots + a_r + a_{r+1}) + (a_{r+2} + a_{r+3} + \cdots + a_s) = \cdots \\
 &= a_1 + a_2 + \cdots + a_r + a_{r+1} + a_{r+2} + \cdots + a_s = \sum_{k=1}^s a_k
 \end{aligned}$$

completing the generalization.

This result may be further generalized through **R5**, viz.,

$$\left(\sum_{i=1}^r a_i \right) + \left(\sum_{j=r+1}^s a_j \right) = \sum_{k=1}^s a_k = \sum_{n=1}^s a_{k_n}$$

i.e.,

$$(a_1 + a_2 + \cdots + a_r) + (a_{r+1} + a_{r+2} + \cdots + a_s) = a_1 + a_2 + \cdots + a_s = a_{k_1} + a_{k_2} + \cdots + a_{k_s}$$

where $a_{k_1}, a_{k_2}, \dots, a_{k_s}$ represent a_1, a_2, \dots, a_s in any order.

- (ii) Likewise
- $\left(\prod_{i=1}^r a_i \right) \left(\prod_{j=r+1}^s a_j \right) = \prod_{k=1}^s a_k$
- , and more generally, i.e. if
- R
- is a commutative ring,

$$\left(\prod_{i=1}^r a_i \right) \left(\prod_{j=r+1}^s a_j \right) = \prod_{n=1}^s a_{k_n}$$

- (iii) Likewise
- $a(b_1 + b_2 + \cdots + b_r) = a \left(\sum_{i=1}^r b_i \right)$
- and
- $(a_1 + a_2 + \cdots + a_r)b = \left(\sum_{i=1}^r a_i \right)b$
- , and more generally,

$$\begin{aligned}
 \left(\sum_{i=1}^r a_i \right) \left(\sum_{j=1}^s b_j \right) &= (a_1 + a_2 + \cdots + a_r)(b_1 + b_2 + \cdots + b_s) \\
 &= a_1 b_1 + \cdots + a_1 b_s + a_2 b_1 + \cdots + a_2 b_s + \cdots + a_r b_1 + \cdots + a_r b_s = \sum_{i=1}^r \sum_{j=1}^s a_i b_j
 \end{aligned}$$

20. Prove Th. 4.1.1.15.

PROOF:

- (i) Define, as is quite customary, $a^1 = a$ and $a^2 = aa = a^{1+1}$; then, in general, $a^m = a^{m-1}a = a^{(m-1)+1}$. Now suppose $a^m a^k = a^{m+k}$; then $a^{m+k}a = a^{(m+k)+1} = a^{m+(k+1)}$. Hence, in general, $a^m a^n = a^{m+n}$.
- (ii) Since, by (i), $(a^m)^1 = a^m$ and $(a^m)^2 = (a^m)(a^m) = a^{m+m} = a^{2m}$, suppose $(a^m)^k = a^{mk}$; then, again by (i), $(a^m)^{k+1} = (a^m)^k(a^m) = a^{mk}a^m = a^{mk+m} = a^{m(k+1)}$. Hence, in general, $(a^m)^n = a^{mn}$.
- (iii) If $ab = ba$, then $(ab)^1 = a^1 b^1$ and $(ab)^2 = (ab)(ab) = aab b = a^2 b^2$, by (i). Hence assume $(ab)^k = a^k b^k$, which then implies $(ab)^{k+1} = (ab)^k(ab) = a^k b^k ab = a^k ab^k b = a^{k+1} b^{k+1}$, and in general, $(ab)^n = a^n b^n$.

21. Prove Th. 4.1.1.16.

PROOF:

- (i) Define, as is exactly the case in elementary algebra, that $1a = a$ and $2a = a+a$, etc; then $1(a+b) = a+b$ and $2(a+b) = (a+b) + (a+b) = a+a+b+b = 2a+2b$, by **R2** and **R5**. Hence, assuming $k(a+b) = ka+kb$, which implies $(k+1)(a+b) = k(a+b) + (a+b) = ka+kb+a+b = ka+a+kb+b = (k+1)a + (k+1)b$, it follows that, in general, $n(a+b) = na+nb$.
- (ii) Since, by (i), $(m+1)a = ma+a$, assume $(m+k)a = ma+ka$ which implies $(m+(k+1))a = (m+k)a+a = ma+ka+a = ma+(k+k)a$. Hence, in general, $(m+n)a = ma+na$.
- (iii) Since $(1a)b = 1ab$ and $(2a)b = (a+a)b = ab+ab = 2ab$, assume $(ka)b = kab$ which then implies $((k+1)a)b = (ka+a)b = kab+ab = (k+1)ab$. Hence, in general, $(na)b = nab$. Likewise, $a(nb) = nab$. Hence, altogether, $(na)b = a(nb) = nab$.

Note. $m(na) = (mn)a$ and $(ma)(nb) = (mn)ab$ can be deduced likewise, using some of (i), (ii), (iii).

§4.1.2 Commutative Rings

4.1.2.1 BOOLEAN RINGS

Df. 4.1.2.1.1 A ring \bar{B} (vs. B , a Boolean algebra, cf. §2.4.2) is called *Boolean* if all of its elements are idempotent, i.e. $a \cdot a = a$ for every element $a \in \bar{B}$.

Stated in weak postulates: a ring is Boolean if it satisfies the following nine axioms, i.e.,

$\bar{B}1-8 = R1-8$.

$\bar{B}9$: Idempotency. $a \cdot a = a$ for every $a \in \bar{B}$.

The following property of \bar{B} , which is deducible from $\bar{B}1-9$, also characterizes \bar{B} , viz.:

Th. 4.1.2.1.2 Every element in \bar{B} is its own additive inverse; i.e. $a \in \bar{B}$ implies $a + a = 0$. (Cf. Prob. 1.)

Th. 4.1.2.1.3 \bar{B} is a class of all subsets of any set S . (Cf. Prob. 2.)

It is due to this theorem that a Boolean ring is sometimes called a *ring of all subsets of a set*.

Th. 4.1.2.1.4 \bar{B} is a commutative ring. (Cf. Prob. 5.)

Th. 4.1.2.1.5 A Boolean ring \bar{B} of two elements 0 and a is isomorphic to the two-value logic L_2 under \vee (complete and exclusive disjunction) and \wedge (conjunction). (Cf. Prob. 6.)

This theorem clarifies the relation between a Boolean ring and a two-value logic, just as the following theorem articulates the difference between a Boolean ring and a Boolean algebra.

Th. 4.1.2.1.6 \bar{B} with unity is a Boolean algebra. (Cf. Prob. 7.)

\bar{B} is also called a *2-ring* in the sense that a *p-ring* is defined as follows:

Df. 4.1.2.1.7 A ring P is a *p-ring* if $a^p = a$ and $pa = 0$ for every $a \in P$.

P is necessarily commutative and, in terms of the *p-ring*, \bar{B} is obviously 2-ring, since $a^2 = a$, by $\bar{B}9$, and $2a = 0$, by Th. 4.1.2.1.2.

Solved Problems

1. Prove Th. 4.1.2.1.2.

PROOF:

Applying $\bar{B}9$ and $\bar{B}8$ repeatedly,

$$\begin{aligned} a + a &= (a + a)(a + a) = a(a + a) + a(a + a) = (aa + aa) + (aa + aa) \\ &= (a + a) + (a + a) \end{aligned}$$

Hence $a + a = (a + a) + (a + a)$, i.e. $0 + (a + a) = (a + a) + (a + a)$, and by Th. 4.1.1.12, $a + a = 0$.

2. Prove Th. 4.1.2.1.3.

PROOF:

Let S be any set and \mathbf{S} be the class of all subsets, including the empty set \emptyset and the universal set U , of S , and denote the elements of \mathbf{S} , which are sets, by A, B, C , etc; then, by the definition of the operations on sets (cf. Df. 2.3.1-3),

$$\mathbf{R1}: A, B \subset \mathbf{S} \text{ implies } A \cup B \subset \mathbf{S}.$$

$$\mathbf{R7}: A \cap (B \cap C) = (A \cap B) \cap C$$

$$\mathbf{R2}: A \cup (B \cup C) = (A \cup B) \cup C$$

$$\mathbf{R8}: A \cap (B \cup C) = (A \cap B) \cup (A \cap C),$$

$$\mathbf{R3}: A \cup \emptyset = \emptyset \cup A = A$$

$$(B \cup C) \cap A = (B \cap A) \cup (C \cap A).$$

$$\mathbf{R4}: A \cup A' = A' \cup A = \emptyset$$

And furthermore,

$$\mathbf{R5}: A \cup B = B \cup A$$

$$\mathbf{R6}: A, B \subset \mathbf{S} \text{ implies } A \cap B \subset \mathbf{S}.$$

$$\bar{\mathbf{B}}9: A \cap A = A.$$

Hence \mathbf{S} , the class of all subsets of any set S , is a Boolean ring, completing the proof.

3. If $a, b \in R$ and $a + b = 0$, then $a = b$.

PROOF:

Since, by Prob. 1 above, $a + a = 0$ and also, by hypothesis, $a + b = 0$, it follows at once that $a + a = a + b = 0$. Then, by Th. 4.1.1.12, $a = b$.

4. Prove that $a + b = a - b$ if $a, b \in B$, and that $a = c + b$ if $a + b = c$ and $a, b, c \in \bar{B}$.

PROOF:

- (i) Since, by Prob. 1, $b + b = 0$ and, by $\mathbf{R4}$, $b + (-b) = 0$, it immediately follows that $b + b = b + (-b) = 0$ and that, by Th. 4.1.1.12, $b = (-b)$. Hence $a + b = a + (-b) = a - b$.
- (ii) $(a + b) + (-b) = c + (-b)$ since $a + b = c$ by hypothesis. But $(a + b) + (-b) = a + (b + (-b)) = a + 0 = a$ and $c + (-b) = c - b = c + b$ by (i). Hence $a = c + b$. (Or, more simply, $c + b = (a + b) + b = a(b + b) = a + 0 = a$, by Prob. 1, i.e. $a = c + b$ if $a + b = c$.)

5. Prove Th. 4.1.2.1.4.

PROOF:

Applying $\bar{\mathbf{B}}9$ and $\mathbf{R5}$ twice and $\mathbf{R2-5}$,

$$\begin{aligned} a + b &= (a + b)(a + b) = a(a + b) + b(a + b) = (aa + ab) + (ba + bb) \\ &= (a + ab) + (ba + b) = (a + b) + (ab + ba) \end{aligned}$$

Hence, by Th. 4.1.1.12, $ab + ba = 0$ and, by Prob. 3, $ab = ba$.

6. Prove Th. 4.1.2.1.5.

PROOF:

Since $0 + 0 = 0$, $a + 0 = 0 + a = a$, by $\mathbf{R3}$; $a + a = 0$, by Th. 4.1.2.4; $0 \cdot 0 = 0$, $a \cdot a = a$, by $\bar{\mathbf{B}}9$; and $a \cdot 0 = 0 \cdot a = 0$, by Th. 4.1.1.13, the following two tables are immediately obtained:

+	0	a
0	0	a
a	a	0

•	0	a
0	0	0
a	0	a

which are indeed isomorphic to:

∨	0	1
0	0	1
1	1	0

∧	0	1
0	0	0
1	0	1

where “0” and “1” denote “false” and “true” respectively.

7. Prove Th. 4.1.2.1.6.

PROOF:

Define $a \cup b = a + b - ab$ and $a \cap b = ab$; then $\mathbf{B1}$ follows at once. Furthermore,

$$\begin{aligned} \mathbf{B2}: (2a) \quad a \cup b &= a + b - ab = b + a - ba = b \cup a, \quad \text{and} \\ (2b) \quad a \cap b &= ab = ba = b \cap a. \end{aligned}$$

B3: (3a) $a \cup (b \cap c) = a + bc - abc.$

$$\begin{aligned} (a \cup b) \cap (a \cup c) &= (a + b - ab)(a + c - ac) = aa + ac - aac + ba + bc - abc - aab - abc + abac \\ &= a + ac - ac + ba + bc - abc - ab - abc + abc && \text{by } \bar{\mathbf{B}}9 \\ &= a + (ac - ac) + (ab - ab) + bc - abc + (abc - abc) && \text{by } \bar{\mathbf{B}}2 \text{ and Th. 4.1.2.5} \\ &= a + 0 + 0 + bc - abc + 0 && \text{by Th. 4.1.2.4 and Prob. 4.} \\ &= a + bc - abc. \end{aligned}$$

$\therefore a \cup (b \cap c) = (a \cup b) \cap (a \cup c).$ And likewise:

$$(3b) \quad a \cap (b \cup c) = (a \cap b) \cup (a \cap c)$$

B4: Define $0 = \emptyset$ and $1 = U$, by hypothesis, and

$$(4a) \quad a \cup 0 = a + 0 - a0 = a, \quad \text{i.e. } a \cup 0 = a, \quad \text{and}$$

$$(4b) \quad a \cap 1 = a1 = a, \quad \text{i.e. } a \cap 1 = a.$$

B5: Define $a' = 1 - a$, and

$$(5a) \quad a \cup a' = a + (1 - a) - a(1 - a) = a + 1 - a - a + aa = (a - a) + 1 + (a - a) = 1, \\ \text{i.e. } a \cup a' = 1, \quad \text{and}$$

$$(5b) \quad a \cap a' = a(1 - a) = a - aa = a - a = 0, \quad \text{i.e. } a \cap a' = 0.$$

B6 follows from **B5**, completing the proof.

8. A Boolean ring \bar{B} with more than two elements is a ring with divisors of zero.

PROOF:

Since, by hypothesis, \bar{B} contains a and b which are distinct and $a \neq 0$ and $b \neq 0$, it also contains, by $\bar{\mathbf{B}}1, 6$, $a + b \neq 0$ (since $a = b$ if $a + b = 0$, by Prob. 3) and ab ; but, by $\bar{\mathbf{B}}8, 7, 9$ and Prob. 2,

$$ab(a + b) = (ab)a + (ab)b = (aa)b + a(bb) = ab + ab = 0$$

That is, ab and $a + b$ are divisors of zero in \bar{B} if $ab \neq 0$. If $ab = 0$, then a and b are divisors of zero themselves, completing the proof.

4.1.2.2 INTEGRAL DOMAINS

Df. 4.1.2.2.1 A ring D is an *integral domain* (or a *domain of integrity*) if it is commutative, with unity, and without zero divisors.

Stated in detail, D satisfies three more axioms in addition to the eight fundamental axioms of the ring in general, viz.:

D1-8 \equiv **R1-8**.

D9: Multiplicative commutativity. $ab = ba$ for every $a, b \in D$.

D10: Multiplicative identity, i.e. unity e (which may be denoted by 1). $ea = ae = a$ (or $1a = a1 = a$) for every $a \in D$.

D11: Multiplicative cancellation. $ac = bc$ and $c \neq 0$ imply $a = b$ for every $a, b, c \in D$.

Note that **D11** implies the nonexistence of zero-divisors in D and conversely (cf. Prob. 1), and that the additive cancellation which has already been proved for R (cf. §4.1.1, Prob. 12) must, of course, hold here.

Example:

The set of all integers (or rational or real or complex numbers) constitutes an integral domain, satisfying all of **D1-11**, but the set of even integers does not, failing to satisfy **D10**, for instance, although it does form a ring in general; similarly, the set of all continuous functions on the closed interval between 0 and 1 fails to form an integral domain because of its difficulties with **D11**.

(Some early authors, like van der Waerden in his *Moderne Algebra* and Dubreil in his *Algèbre*, defined D without **D10**, their integral domains being merely commutative rings without zero-divisors. In the following pages, however, an integral domain is to satisfy all of **D1-11**, unless otherwise modified.)

Df. 4.1.2.2.1a If a complex D' of an integral domain D forms an integral domain itself, then D' is called a *subdomain* of D .

Example:

The set I of all integers is a subdomain of the set R of all rational numbers, while R is in turn a subdomain of the set R^* of all real numbers. As in each case, a subdomain D' is closed relative to the operations of the given domain D and contains both additive and multiplicative identities and also additive inverses in itself.

Note. I , the set of integers, is sometimes called a *minimal* integral domain as it contains no subdomains (cf. Th. 4.2.1.4).

Th. 4.1.2.2.2 The only idempotents in D are 0 and 1. (Cf. Prob. 2.)

This theorem draws a clear line of demarcation between D and \bar{B} (cf. Df. 4.1.2.1.1); commutative rings as they both are, their difference is quite conspicuous. The difference manifests itself even more clearly through the following definition and theorem.

Df. 4.1.2.2.3 The *characteristic* of a ring R is the smallest positive integer n such that $na = 0$ for every $a \in R$. In particular, R is of *characteristic zero* (or, as is sometimes called, *infinite characteristic*) if $na \neq 0$ for every $a \in R$.

Example:

\bar{B} is obviously of characteristic 2 (cf. Th. 4.1.2.1.2), and the characteristic of the additive identity in R is 1, while every familiar number system, excluding 0, of elementary algebra has characteristic zero (or infinite). Moreover, a ring R of integers modulo m is of characteristic m itself, since the smallest positive integer k such that $ka \equiv 0 \pmod{m}$ for any $a \in R$ is m itself (cf. §4.1.1, Prob. 8-9).

Th. 4.1.2.2.4 The characteristic of every integral domain D is either zero or a prime. (Cf. Prob. 3.)

Residue classes, then, cannot always form integral domains, as is quite clearly stipulated by this theorem, since many residue classes do contain zero-divisors (against D11). In the residue class modulo 6, for instance, $2 \neq 0$ and $3 \neq 0$, yet $2 \cdot 3 = 0$; hence the set of integers modulo 6 is not an integral domain, although the set of integers of 7, for example, is.

Df. 4.1.2.2.5 An integral domain D is called *ordered*, and denoted by \bar{D} , if it contains a complex D^+ whose elements, called the *positive* elements, satisfy the following conditions:

- (i) Closure under addition: $a, b \in D^+$ implies $a + b \in D^+$.
- (ii) Closure under multiplication: $a, b \in D^+$ implies $a \cdot b \in D^+$.
- (iii) Trichotomy: $a \in D^+$ implies one, and only one, of three mutually exclusive alternatives, viz. $a > 0$, reading " a is greater than 0", or $a = 0$, or $-a > 0$.

In this context the additive inverse of a , denoted by $-a$ and called the *negative* or *minus* a (cf. Df. 4.1.1.1), is defined as an element which cannot belong to D^+ if a does, i.e. if a is positive. On the other hand, if a does not belong to D^+ , i.e. if a is not positive, hence negative, then $-a$ is negative negative, hence positive, and belongs to D^+ .

Note. $-a > 0$ may be written as $a < 0$, reading " a is less than 0". This notation may rewrite Df. 4.1.2.2.5 as follows:

- (i-ii) $a > 0$ and $b > 0$ imply $a + b > 0$ and $a \cdot b > 0$.
- (iii) $a > 0$ or $a = 0$ or $a < 0$ for every $a \in D^+$.

The same concept is further amplified and clarified by the following definition.

Df. 4.1.2.2.6 $a > b$ and $b < a$ are logically equivalent in \bar{D} , both meaning the same: $a - b$, which is called the *difference* between a and b , is positive.

If $a - b$ is not positive, then it is either zero or negative, i.e. either $a = b$ or $a < b$; these two alternatives are often combined in one notation, $a \leq b$, just as $a = b$ and $a > b$ may be incorporated into one, $a \geq b$, meaning that $a - b$ is not negative.

Note. $0 = a - a$ for every $a \in \bar{D}$, i.e. zero is defined in terms of difference as the difference between an element of \bar{D} and itself.

Df. 4.1.2.2.7 A complex S , $S \subseteq \bar{D}$, is called *well-ordered* (cf. Df. 2.4.1.18) if any subcomplex R of S contains a least element a such that $a \leq r$ for every $r \in R$.

Th. 4.1.2.2.8 If an ordered integral domain \bar{D}_1 contains a complex D_1^+ of all positive elements of \bar{D}_1 , and if \bar{D}_2 contains a similar complex D_2^+ , then \bar{D}_1 and \bar{D}_2 are isomorphic. (Cf. Prob. 9.)

The elements of \bar{D} , in general, have the following properties:

Th. 4.1.2.2.9 All squares of non-zero elements in \bar{D} are positive. (Cf. Prob. 10.)

Th. 4.1.2.2.10 Transitivity holds in \bar{D} : $a < b$ and $b < c$ imply $a < c$, for every $a, b, c \in \bar{D}$.

Th. 4.1.2.2.11 For every $a, b, c \in \bar{D}$, $a > b$ and $c > 0$ imply $a + c > b + c$ and $ac > bc$. (Cf. Prob. 12.)

In any ordered integral domain, defined and expanded as above, it is already feasible to introduce the concept of absolute values, with which the student is quite familiar, as below.

Df. 4.1.2.2.12 For every element $a \in \bar{D}$, the *absolute value* of a , denoted by $|a|$, is positive except when $a = 0$ for which $|a| = 0$.

This definition yields the following results.

Th. 4.1.2.2.13 For every $a, b \in \bar{D}$,

$$(i) \quad |a + b| \leq |a| + |b| \quad (ii) \quad |ab| = |a||b|$$

(Cf. Prob. 13-14.)

This theorem may be considered a special case of the so-called (i) *triangle inequality* and (ii) *Schwarz inequality*, respectively, the general case of which is expounded in terms of rational and real numbers (cf. §5.1.1, Prob. 13) and also, with a slight modification, in terms of complex numbers (cf. §5.1.3, Prob. 12, 19).

Solved Problems

1. If the cancellation law holds for a ring R , then $a \neq 0 \in R$ is not a zero-divisor, and conversely.

PROOF:

- (i) Let $b \in R$ and $ab = 0$; then $a \cdot b = a \cdot 0 = 0$ since $a \cdot 0 = 0 \cdot a = 0$, by Th. 4.1.1.13, and the cancellation law implies $b = 0$. Likewise $ba = 0$ implies $b = 0$. Hence a is not a zero-divisor.
- (ii) Conversely, if $a \neq 0$ is not a zero-divisor, then $ab = ac$ implies $ab - ac = a(b - c) = 0$, which in turn implies $b - c = 0$ and $b = c$. Likewise $ba = ca$ implies $b = c$, establishing the cancellation law.

Note. This theorem is logically equivalent to the theorem that a product of non-zero factors in D is not zero.

2. Prove Th. 4.1.2.2.2.

PROOF:

- (i) 0 and 1 are evidently idempotents in D , since $0 \cdot 0 = 0$, by Th. 4.1.1.13, and $1 \cdot 1 = 1$, by D10.

- (ii) Let a be neither 0 nor 1, and suppose $a \cdot a = a$. Then, by D10, $a \cdot 1 = 1 \cdot a = a$ and, by the assumption, $a \cdot a = a \cdot 1 = a$. Hence, by D11, $a = 1$, proving that D has neither more nor less than two idempotents.

3. Prove Th. 4.1.2.2.4.

PROOF:

Assume that D is of characteristic n which is not a prime and is also greater than 0, and let $n = xy$, where $1 < x < n$ and $1 < y < n$. If e is the unity of D , then, by Df. 4.1.2.2.3, $ne = 0$, i.e. $(xy)e = 0$, which implies $(xe)(ye) = 0$, which in turn implies, by D11, that either $xe = 0$ or $ye = 0$.

Now assume $xe = 0$; then, for any $a \in D$, $xa = x(ea) = (ex)a = 0$, implying that $xa = 0$ for any element a of D , which is contradictory to the assumption that $1 < x < n$. Hence n must be either zero or a prime.

The same conclusion is obtained by assuming $ye = 0$, completing the proof.

4. The order p , $p > 0$, of the additive cyclic group generated by the unity e of an integral domain D is a prime.

PROOF:

Suppose $p = mn$, where m and n are any two integers; then $me, ne \in D$, and by the definition of cyclic groups, $(me)(ne) = (e + e + \dots + e)(e + e + \dots + e) = (mn)e^2 = (mn)e = 0$.

But, since D is an integral domain which by definition cannot have proper zero-divisors, it must be the case that either $me = 0$ or $ne = 0$, either of which is contradictory to the definition of the order p of the cyclic group generated by e . Hence p must be a prime.

Note that this theorem actually reassures the validity of Th. 4.1.2.2.4 (Prob. 3 above).

5. All non-zero elements of D generate additive cyclic groups of the same order.

PROOF:

Let e and a be the unity element and non-zero element, respectively, of D whose characteristic is, by Th. 4.1.2.2.4, either p or 0.

- (i) If D is of characteristic p , then $pa = p(ae) = p(ea) = (pe)a = 0$, which implies, by D11, that $pe = 0$ and, by Prob. 4 above, that p is the prime order of the additive cyclic group generated by any non-zero element of D .
- (ii) If D is of characteristic zero, then $na \neq 0$ for any $n \neq 0$, which implies that the additive cyclic group generated by a (any non-zero element of D) is of the same infinite order, completing the proof.

6. An integral domain D of characteristic zero contains a complex C which is isomorphic to the integral domain I of integers.

PROOF:

Let every element $c \in C$ ($C \subset D$ by hypothesis) be of the form ne , where e is the unity of D and n is any integer, i.e. $n \in I$; then C forms an infinite additive cyclic group whose elements are all distinct. Hence the 1-1 correspondence $c = ne \in C \leftrightarrow n \in I$ is an isomorphism, since $n_1e \leftrightarrow n_1$ and $n_2e \leftrightarrow n_2$ do imply

$$n_1e + n_2e = (n_1 + n_2)e \leftrightarrow n_1 + n_2 \quad \text{and} \quad (n_1e)(n_2e) = (n_1n_2)e \leftrightarrow n_1n_2$$

7. An element d of an integral domain D both divides the unity e of D and is divisible by e iff its multiplicative inverse d^{-1} is also in D .

PROOF:

If d divides e , then there exists an element $c \in D$ such that $cd = e$, which proves that $c = d^{-1}$.

Conversely, if d has a multiplicative inverse c , then $cd = e$, which proves that e is divisible by d .

8. If a subset \bar{D}^+ of the positive elements of an ordered integral domain \bar{D} is well-ordered, then

$$(i) \quad \bar{D}^+ = \{n_1e\} \quad \text{and} \quad (ii) \quad \bar{D} = \{n_2e\}$$

where e is the unity of \bar{D} , n_1 any positive integer, and n_2 any integer.

PROOF:

- (i) By hypothesis, \bar{D}^+ has a least element, which in this case is e , since the assumption $0 < a < e$, $a \in \bar{D}^+$ implies an immediate contradiction that $a^2 < a$ (since $ae = a$), $a^2 \in \bar{D}^+$.

Furthermore, since $e^2 = e$, by D4, it follows that $e > 0$ and $e^2 > 0$, and that $2e = e + e > 0$, $3e = 2e + e > 0$, and in general $n_1e = (n_1 - 1)e + e > 0$. Hence $n_1e \in \bar{D}^+$ for any positive integer n_1 .

Conversely, every element of \bar{D}^+ is necessarily of this form. For, otherwise, a complex C of \bar{D}^+ , whose elements are assumed to be not of the form, must have a least element, say b . Then $b > e$, i.e. $b - e > 0$, since e has already been proved to be the least element of \bar{D}^+ . Hence $b - e \in \bar{D}^+$ and $b - e < b$ (since $e > 0$), which implies $b - e \notin C$, which in turn implies $b - e = n_1e$, i.e. $b = n_1e + e = (n_1 + 1)e$, where $n_1 + 1$ is of course a positive integer, which is contradictory to the assumption. Hence C must be a null set, proving that every element of \bar{D}^+ must be of the form n_1e .

- (ii) Suppose that $d \in \bar{D}$ and $d \notin \bar{D}^+$; then, by Df. 4.1.2.2.5, either $d = 0$ or $-d \in \bar{D}^+$. In the former case, $d = 0 \cdot e$, and in the latter case, $-d = n_1e$ for any positive integer n_1 , as has just been proved above. In either case $d = n_2e$ for any integer n_2 (zero or positive or negative), completing the proof.

9. Prove Th. 4.1.2.2.8.

PROOF:

Let e_1 and e_2 be the unities of \bar{D}_1 and \bar{D}_2 respectively; then, by Prob. 11, $\bar{D}_1 = S_1$ and $\bar{D}_2 = S_2$, where $S_1 = \{ne_1\}$ and $S_2 = \{ne_2\}$ for any integer n .

Suppose $n_1e_1 = n_2e_1$ when $n_1 \neq n_2$, say $n_1 > n_2$, i.e. $n_1 - n_2 > 0$; then $(n_1 - n_2)e_1 = 0$, which is contradictory to the result of Prob. 11, according to which $(n_1 - n_2)e_1 > 0$. Hence each element of \bar{D}_1 must be uniquely expressible by ne_1 , and every element of \bar{D}_2 by ne_2 .

The uniqueness of each element of \bar{D}_1 and \bar{D}_2 at once entails the distinct elements of respective sets, which then enables a definite 1-1 mapping between them, i.e.,

$$n_1e_1 + n_2e_1 = (n_1 + n_2)e_1 \leftrightarrow (n_1 + n_2)e_2 = n_1e_2 + n_2e_2$$

and

$$(n_1e_1)(n_2e_1) = (n_1n_2)e_1 \leftrightarrow (n_1n_2)e_2 = (n_1e_2)(n_2e_2)$$

completing the proof.

10. All squares of non-zero elements in an ordered domain \bar{D} are positive.

PROOF:

Let $a \in \bar{D}$, $a \neq 0$; then, by Df. 4.1.2.2.5,iii, either a or $-a$ is positive. Hence, in the first case, $a \cdot a = a^2 \in D^+$, by Df. 4.1.2.2.5,ii, and in the second case, $(-a)(-a) = a \cdot a = a^2 \in D^+$, by Th. 4.1.1.14,v (which of course holds for \bar{D}), completing the proof.

11. Prove Th. 4.1.2.2.10.

PROOF:

The hypotheses $a < b$ and $b < c$ are logically equivalent, by Df. 4.1.2.2.6, to $b - a > 0$ and $c - b > 0$, and $(b - a) + (c - b) > 0$, by Df. 4.1.2.2.5,i. But $(b - a) + (c - b) = (b - a) + (c - a) = c - a$. Hence $c - a > 0$, i.e. $a < c$, completing the proof.

12. Prove Th. 4.1.2.2.11.

PROOF:

- (i) $(a + c) - (b + c) = (a - b) + (c - c) = a - b > 0$ since, by hypothesis, $a > b$, i.e. $a - b > 0$. Hence $a + c > b + c$.
- (ii) $ac - bc = (a - b)c > 0$ since, by hypothesis, $a - b > 0$, $c > 0$, and consequently, by Df. 4.1.2.2.5,ii, $(a - b)c > 0$. Hence $ac > bc$.

13. Prove that $|a + b| \leq |a| + |b|$ if $a, b \in \bar{D}$.

PROOF:

- (i) If $a \geq 0$ and $b \geq 0$, then $a + b \geq 0$ and

$$|a + b| = a + b = |a| + |b|$$

- (ii) If $a \leq 0$ and $b \leq 0$, then $-a \geq 0$, $-b \geq 0$, $-(a + b) = (-a) + (-b) \geq 0$, and

$$|a + b| = -(a + b) = (-a) + (-b) = |a| + |b|$$

(iii) If $a > 0$ and $b < 0$, then $b < -b$, which implies

$$a + b < a + (-b) = |a| + |b| \quad \text{and} \quad -(a + b) = (-a) + (-b) < a + (-b) = |a| + |b|$$

Hence

$$|a + b| < |a| + |b|$$

(iv) Similarly, if $a < 0$ and $b > 0$, then

$$|a + b| < |a| + |b|$$

Hence, from (i)-(iv), which exhaust all possible cases,

$$|a + b| \leq |a| + |b|$$

Note. This result can be readily generalized, by mathematical induction, to:

$$\left| \sum_{i=1}^n a_i \right| \leq \sum_{i=1}^n |a_i|, \quad \text{if } a_1, a_2, \dots, a_n \in \bar{D}$$

14. Prove that $a, b \in \bar{D}$ implies $|ab| = |a||b|$.

PROOF:

(i) If $a > 0$ and $b > 0$, then $ab > 0$; hence

$$|ab| = ab = |a||b|$$

(ii) If $a < 0$ and $b < 0$, then $-a > 0$, $-b > 0$, $(-a)(-b) > 0$, and

$$|ab| = |(-a)(-b)| = (-a)(-b) = |a||b|$$

(iii) If $a > 0$ and $b < 0$, then $-b > 0$, $a(-b) > 0$, and

$$|ab| = |-ab| = |a(-b)| = a(-b) = |a||b|$$

(iv) Likewise, if $a < 0$ and $b > 0$, then

$$|ab| = |a||b|$$

Hence, from (i)-(iv), which exhaust all possible cases,

$$|ab| = |a||b|$$

Note. This result, just as in Prob. 8, can be generalized, by mathematical induction, to:

$$\left| \prod_{i=1}^n a_i \right| = \prod_{i=1}^n |a_i|, \quad \text{if } a_1, a_2, \dots, a_n \in \bar{D}$$

15. Prove that $a, b \in \bar{D}$ implies $||a| - |b|| \leq |a \pm b| \leq |a| + |b|$.

PROOF:

Since $a = (a - b) + b$ and $b = (b - a) + a$, it follows, from Prob. 13, that

$$|a| \leq |a - b| + |b| \quad \text{and} \quad |b| \leq |b - a| + |a|$$

Hence

$$|a| - |b| \leq |a - b| \quad \text{and} \quad |b| - |a| \leq |a - b|$$

i.e.

$$||a| - |b|| \leq |a - b| = |a + (-b)| \leq |a| + |b|$$

But $a + b = a - (-b)$ and $|b| = |-b|$. Hence

$$||a| - |b|| \leq |a \pm b| \leq |a| + |b|$$

4.1.2.3 INTEGERS

Df. 4.1.2.3.1 (Peano Axioms). The set N of *natural numbers* satisfies the following four axioms:

- N1. $a \leftrightarrow S(a) = a'$ in N is a 1-1 correspondence for every $a \in N$ such that $a' \in N$, where the mapping $S(a)$ is called the *successor function* and a' the *successor* of a .
- N2. $1 \in N$ and, for every $a \in N$, $S(a) \neq 1$; i.e. 1 is a natural number, yet never the successor of any natural number.
- N3. $S(a) = S(b)$ iff $a = b$ for every $a, b \in N$.
- N4. $N = M$ if a complex M of N contains 1 and if $a \in M$ implies $S(a) \in M$.

The undefined terms (i.e. primitive terms, cf. Df. 2.1.1) in this definition are “natural numbers”, “1”, and “successor”. It should be noted that **N4** is in essence the Principle of Finite Induction (cf. MTh. 2.2.1.11), and that there exist postulate sets other than Peano’s, e.g. von Neumann’s, for defining natural numbers.

Df. 4.1.2.3.2 The binary operation in N under $+$, defined by

$$(i) \quad a + 1 = a', \quad (ii) \quad a + b' = (a + b)'$$

for every $a, b \in N$, is called *addition*.

In terms of this definition, **N1** can be simply replaced by $S(a) = a + 1$ or $a' = a + 1$. This operation is evidently closed in N , since every element of N except 1 is of the form a' ; it is also associative, commutative, allowing cancellation under addition (cf. Prob. 1-3).

Df. 4.1.2.3.3 The binary operation in N under \cdot , defined by

$$(i) \quad a \cdot 1 = a, \quad (ii) \quad a \cdot b' = ab + a$$

for every $a, b \in N$, is called *multiplication*.

Multiplication is obviously closed in N , since every element of N , except 1, is the successor of some element; it is also associative, commutative, distributive under addition (cf. Prob. 6-8).

It is clear, then, that the set N of natural numbers satisfies every property of an integral domain except **D3** (additive identity) and **D4** (additive inverse), including Df. 4.1.2.2.5; N , then, is an ordered set (cf. Prob. 9-14), although it does not form even a group, much less a ring and still less an integral domain. Conversely, however:

Th. 4.1.2.3.4 Every ordered integral domain \bar{D} contains a unique complex \bar{N} of positive elements which satisfies Df. 4.1.2.3.1. (Cf. Prob. 17.)

The set N of natural numbers being thus defined, the set J (or I) of all integers is developed from N in the direction suggested by Df. 4.1.2.2.5, viz. to introduce all integers as *ordered pairs* (a, b) of natural numbers a and b ; for the *difference* $a - b$ can be either positive or zero or negative.

Df. 4.1.2.3.5 All ordered pairs of the form (x, y) , where $x, y \in N$, are called *integers*, forming the set J of integers, in which $(a, b) = (c, d)$ iff $(a, b), (c, d) \in J$ and $a - b = c - d$, i.e. $a + d = b + c$.

An element of J , defined as above, which is ordinarily called an integer, may be sometimes called a *rational integer* to distinguish it from the specific set \bar{I} of *algebraic integers* (cf. Df. 5.3.2.13). Note, also, that the equality in Df. 4.1.2.3.5 is an equivalence relation (cf. Prob. 18).

Df. 4.1.2.3.6 The binary operations of addition and multiplication in J are, respectively:

$$(i) \quad (a, b) + (c, d) = (a + c, b + d), \quad (ii) \quad (a, b) \cdot (c, d) = (ac + bd, ad + bc)$$

for every $(a, b), (c, d) \in J$.

Th. 4.1.2.3.7 Addition in J is associative, commutative, and well-defined; so is multiplication in J , which is also distributive under addition. (Cf. Prob. 20-22.)

The set J of integers as such is manifestly an integral domain, satisfying **D1-11**; as has already been observed (cf. Df. 4.1.2.2.1a), then, J exemplifies an ordered integral domain, and also, as has been revealed by Th. 4.1.2.2.4, embodies a finite integral domain through the ring I_p of integers modulo p , a prime.

Th. 4.1.2.3.8 The set N of natural numbers is isomorphic to the set J^+ of positive integers under addition, multiplication, and order. (Cf. Prob. 23.)

Such an isomorphism as above, which preserves order, is specifically called an *order-isomorphism* between two ordered sets, although it may not be explicitly mentioned as such (cf. Th. 4.2.1.3-7).

Since I (or J), or more generally D or \bar{D} which contains I , does not always assure an integral solution x for $ax = b$, where $a, b \in I$, division is evidently restricted in I , as is clearly reflected in the familiar term “*integral operation*” which includes addition, subtraction, and multiplication, but excludes division. Here is divisibility in general (as in “*rational operation*”) at issue:

Df. 4.1.2.3.9 If there exists an element $x \in I$ for any $c, d \in I$ such that

$$cx = d \quad (1)$$

then d is said to be *divisible* by c , where d is called a *multiple* of c , which in turn is called a *divisor* or *factor* of d . The equation (1) may be expressed by $c|d$, which reads: c divides d . Similarly, $c \nmid d$ denotes $cx \neq d$, reading: c does not divide d . E.g. $2|4$ and $2 \nmid 5$.

Note. 0 is considered divisible by every element of I or D , since

$$a \cdot 0 = 0$$

for every element $a \in I$ or $a \in D$.

Df. 4.1.2.3.10 If $c \in I$ is neither 0 nor ± 1 , and if there exists $x \in I$, also neither 0 nor ± 1 , such that $cx = d$, where $d \in I$, then c is called a *proper divisor* of d . A *prime* is then defined as an integer which is neither 0 nor ± 1 and has no proper divisors.

Stated otherwise, a prime p is neither 0 nor ± 1 and divisible only by ± 1 and $\pm p$.

Note. $d \in I$ is sometimes called a *composite (number)* if it has proper divisors.

Th. 4.1.2.3.11 Divisibility is reflexive and transitive, but not always symmetric. (Cf. Prob. 26.)

Df. 4.1.2.3.12 a and b , where $a, b \in I$ or $a, b \in D$, are called *associates* if division is symmetric, i.e. $a|b$ and $b|a$. In particular, an associate of $e \in D$ is called a *unit* (cf. Prob. 27-28 below). In this sense, too, an element $p \in D$ is said to be a *prime element* if its only divisors are units and elements associated with p .

Example:

$2 + \sqrt{3}$ and $2 - \sqrt{3}$ are units of an integral domain whose elements are of the form: $a + b\sqrt{3}$, where $a, b \in I$.

Th. 4.1.2.3.13 (Division Algorithm). Given two positive integers a and b , there always exist two unique non-negative integers q and r such that

$$a = bq + r, \quad 0 \leq r < b$$

(Cf. Prob. 29.)

Example:

$a = 50$ and $b = 11$ imply $q = 4$ and $r = 6$ in the context of Th. 4.1.2.3.13. For the same b , however, $a = -50$ implies $q = -5$, $r = 5$, and likewise $a = -5$ implies $q = -1$ and $r = 6$.

Note. q above is called a *quotient* and r a *remainder* – terms with which the student is quite familiar, just as with the following terms:

Df. 4.1.2.3.14 If $a|m$ and $b|m$, where $a, b, m \in I$, then m is called a *common multiple* of a and b ; and if furthermore $a|n$ and $b|n$ imply $m|n$, where $n \in I$, then m is called a *least common multiple (l.c.m.)* of a and b , denoted by $m = [a, b]$.

Similarly, if $d|a$ and $d|b$, where $a, b, d \in I$, then d is called a *common divisor* of a and b ; and if moreover $c|a$ and $c|b$ imply $c|d$, where $c \in I$, then d is called a *greatest common divisor (g.c.d.)* of a and b , denoted by $d = (a, b)$.

Example:

$$6 = [2, 3] \quad \text{and} \quad 3 = (6, 9).$$

Df. 4.1.2.3.15 If in particular $(a, b) = 1$, then a and b are called *relatively prime*; i.e. a and b have no common divisors except ± 1 , while 1 is relatively prime to every integer including itself. (Cf. Prob. 33-34.)

Furthermore, any two elements $a, b \in I$, $a \neq 0$, $b \neq 0$, have a l.c.m. $[a, b]$, which is always obtainable by the so-called *Euclidean Algorithm* (cf. Prob. 31), and have also a g.c.d. (a, b) , viz.:

Th. 4.1.2.3.16 Any two nonzero elements $a, b \in I$ have a g.c.d., which is always expressible in the form:

$$ax + by = (a, b), \quad x, y \in I,$$

which in turn is called a *linear combination* of a and b . (Cf. Prob. 32.)

Also, as has already been tacitly presumed, any nonzero integer can be factored, in fact uniquely, as is revealed in the following theorem of unique factorization, known as the *fundamental theorem of arithmetic*, viz.:

Th. 4.1.2.3.17 Any positive integer n can be expressed as a product of positive primes, and this expression, except for the order of the factors, is unique. (Cf. Prob. 36.)

The restriction to positive integers is merely for the sake of simplicity in proof, as is evident in the context.

Since the set I of integers is an integral domain, which in turn is *a fortiori* a module, i.e. an Abelian group under addition (cf. D1-5 and also the resemblance between Df. 4.2.6.6a and Df. 4.1.2.2.6), it may form residue classes in accordance with the following definition:

Df. 4.1.2.3.18 If $m|(a - b)$, $a, b, m \in I$, i.e. if there exists an element k , $k \in I$, such that $a - b = km$, then a is said to be *congruent* to b *modulo* m ; notationally,

$$a \equiv b \pmod{m}$$

where m is the *modulus* of the congruence. The residue classes of integers modulo m may be denoted simply by I_m , instead of $I/\{m\}$, $I/(m)$, etc., if there exists no danger of misinterpreting subscripts.

Conversely, as has been proved by Th. 4.1.2.2.4, I_m is an integral domain iff m is a prime. In either case, due to the additional axioms of D6-11, congruence operations in I_m (or more generally in D if m is a prime) is more complex than in a plain module, as is evident in the following theorems.

Th. 4.1.2.3.19 If $a \equiv b \pmod{m}$, then, for any $x \in I$,

$$(i) \quad a + x \equiv b + x \pmod{m}, \quad (ii) \quad ax \equiv bx \pmod{m}$$

(Cf. Prob. 37.)

Th. 4.1.2.3.20 The congruence $ax \equiv b \pmod{m}$ has an integral solution x iff $d \mid b$, where $d = (a, m)$; then, if it does, there uniquely exists a representative set of d solutions modulo m . (Cf. Prob. 41.)

Congruences can be treated simultaneously, viz.:

Th. 4.1.2.3.21 The congruences $x \equiv a_1 \pmod{m_1}$ and $x \equiv a_2 \pmod{m_2}$ have a common solution iff $a_1 \equiv a_2 \pmod{(m_1, m_2)}$ and, then, there is a single solution for the modulus $[m_1, m_2]$. (Cf. Prob. 42.)

This theorem may be further generalized to a set of n simultaneous congruences (cf. Prob. 43).

Solved Problems

1. Addition in N is associative.

PROOF:

Let M be a complex of N , for which $a + (b + c) = (a + b) + c$ holds for any c , given a and b . Then, by Df. 4.1.2.3.2,

$$(a + b) + 1 = (a + b)' = a + b' = a + (b + 1)$$

which implies $1 \in M$. Furthermore, by Df. 4.1.2.3.2,

$$(a + b) + c' = ((a + b) + c)' = (a + (b + c))' = a + (b + c)' = a + (b + c')$$

which implies $c' \in M$ and, by N4 of Df. 4.1.2.3.1, establishes

$$a + (b + c) = (a + b) + c$$

for N itself.

Note. The proof has been completed, in effect, by justifying induction for any c and fixed a and b ; this procedure will be seen again in the following problems.

2. Multiplication in N is commutative.

PROOF:

- (i) Let M be a complex of N and $a \in M$, for which $a + 1 = 1 + a$, which implies $1 \in M$ and $a' \in M$, since

$$a' + 1 = (a + 1) + 1 = (1 + a) + 1 = (1 + a)' = 1 + a'$$

Hence, by N4, $a + b = b + a$ holds for $b = 1$.

- (ii) Let M be a complex of N , where $a + b = b + a$ holds for b , given a , and, by the result of (i) above, $1 \in M$; then

$$a + b' = (a + b)' = (b + a)' = b + a' = b + (a + 1) = b + (1 + a) = (b + 1) + a = b' + a$$

proving $b' \in M$. Hence, by N4, $a + b = b + a$ holds in N itself, completing the proof.

3. For every $a, b, c \in N$, $b \neq c$ implies $a + b \neq a + c$.

PROOF:

Let M be a complex of N , in which $a + b \neq a + c$ holds for a , given b and c . Then, since $b \neq c$ implies $1 + b \neq 1 + c$, by N3, it follows that $1 \in M$.

Now, in general, suppose $a \in M$ and $b \neq c$ imply $a + b \neq a + c$. Then, by the result obtained at the start, $(a + b)' \neq (a + c)'$, i.e. $a' + b \neq a' + c$, proving that $a' \in M$ and that, by N4, the theorem holds for N itself.

Note. The theorem may be stated otherwise, by MTh. 1.1.1.12, that $a + b = a + c$ implies $b = c$ for every $a, b, c \in N$, which is a more direct statement of cancellation under addition.

4. $a + b \neq b$ for every $a, b \in N$.

PROOF:

The theorem evidently holds for $b = 1$, since $a + 1 \neq 1$ by N2 of Df. 4.1.2.3.1.

In general, $a + b \neq b$ implies $a + b' = (a + b)' \neq b$, by N3 (or rather its contrapositive: $a \neq b$ iff $a' \neq b'$), proving the case for b' . Hence, by N4, the theorem holds for every $a, b \in N$.

5. One, and only one, of the following three cases is possible for any $a, b \in N$:

(i) $a = b$, or (ii) $a = b + m$ for some $m \in N$, or (iii) $b = a + n$ for some $n \in N$

PROOF:

It immediately follows from Prob. 4 that neither (i) and (ii) nor (i) and (iii) can hold simultaneously.

Now, assume that (ii) and (iii) hold simultaneously; then

$$a = b + m = (a + n) + m = a + (m + n)$$

which is contradictory to Prob. 4. Hence (i), (ii), and (iii) cannot hold simultaneously. One of the three, however, always holds for the following reason.

Let M be a complex of N for which one of (i), (ii), (iii) holds for b , given a ; then (i) holds if $a = 1$ and $b = 1$, and (ii) holds if $a \neq 1$ and $b = 1$, since, by N2, $a = m' = m + 1 = 1 + m$. Hence $1 \in M$.

In general, $b \in M$ implies $a = b$ or $a = b + m$ or $b = a + n$.

In the first case, i.e. if $a = b$, then $b' = b + 1 = a + 1$, which implies (iii) holds for b' .

In the second case, i.e. if $a = b + m$, then $a = b + 1 = b'$ for $m = 1$, which implies that (i) holds for b' . For $m \neq 1$, i.e. $m = k'$, it follows that $a = b + k' = b + (k + 1) = b + (1 + k) = (b + 1) + k = b' + k$, establishing (ii) for b' .

In the third case, i.e. if $b = a + n$, it follows likewise that $b' = (a + n)' = a + n'$, establishing (iii) for b' , which completes the proof.

6. For every $a, b, c \in N$, $(a + b)c = ab + bc$ and $a(b + c) = ab + ac$.

PROOF:

(i) The theorem holds for $c = 1$, given a and b , since $(a + b) \cdot 1 = a + b = a \cdot 1 + b \cdot 1$.

In general, if the theorem holds for c , given a and b , then it holds also for c' , since, by Prob. 2-3,

$$(a + b)c' = (a + b)c + (a + b) = (ac + bc) + (a + b) = (ac + a) + (bc + b) = ac' + bc'$$

Hence, by N4, $(a + b)c = a + c + bc$ for every $a, b, c \in N$.

(ii) Likewise, $a(b + c) = ab + ac$.

7. Multiplication in N is commutative, i.e. $ab = ba$ for every $a, b \in N$.

PROOF:

(i) Let M be a complex of N , for which $1 \cdot b = b \cdot 1$ holds, implying $1 \in M$; then, since $1 \cdot b = b \cdot 1$ implies

$$1 \cdot b' = 1 \cdot b + 1 = b \cdot 1 + 1 = b + 1 = b' = b' \cdot 1$$

it follows that $b' \in M$, and that, by N4, the theorem generally holds for b , given $a = 1$.

(ii) Let M be now a complex of N where $a \cdot b = b \cdot a$ holds for b , given a ; then $1 \in M$, by (i) above, and $a \cdot b = b \cdot a$ implies, by Prob. 6,

$$ab' = a(b + 1) = ab + a \cdot 1 = ba + a \cdot 1 = ba + a = (b + 1)a = b'a$$

establishing $b' \in M$, which in turn proves, by N4, commutativity in general in N .

8. Multiplication in N is associative, i.e. $a(bc) = (ab)c$ for every $a, b, c \in N$.

PROOF:

If M is a complex of N in which the theorem holds for c , given a and b , then $1 \in M$, since $a(b \cdot 1) = ab = (ab) \cdot 1$.

In general, $a(bc) = (ab)c$ in M implies

$$a(bc') = a(bc + b) = a(bc) + ab = (ab)c + ab = (ab)c'$$

proving $c' \in M$, which in turn proves, by N4, associativity in N .

9. One, and only one, of the following three alternatives holds for every $a, b \in N$:

$$a > b \quad \text{or} \quad a = b \quad \text{or} \quad a < b$$

PROOF:

This is merely a restatement of Prob. 5, since $a > b$ iff $a = b + m$ for some $m \in N$, or $a < b$ iff $b = a + n$ for some $n \in N$.

10. For every $a, b, c \in N$, $a < b$ and $b < c$ imply $a < c$.

PROOF:

Since, by Prob. 5, $a < b$ and $b < c$ imply $b = a + m$ and $c = b + n$ for some $m, n \in N$, it follows at once that $c = b + n = (a + m) + n = a + (m + n)$, which implies $a < c$, completing the proof.

11. For every $a, b, c \in N$, $a + c < b + c$ and $ac < bc$ iff $a < b$. The dual also holds, i.e. $a + c > b + c$ and $ac > bc$ iff $a > b$.

PROOF:

- (i) If $a < b$, then $b = a + m$ for some $m \in N$, which implies

$$b + c = (a + m) + c = a + m + c = a + c + m = (a + c) + m \quad \text{and} \quad bc = (a + m)c = ac + cm$$

which in turn immediately implies $a + c < b + c$ and $ac < bc$ respectively.

The converse follows, since, as in Prob. 5, $b + c = (a + c) + m$ for some $m \in N$, i.e. $b + c = (a + m) + c$, which implies, by Prob. 3, $b = a + m$, i.e. $a < b$. Likewise, $ac < bc$ implies $a < b$.

- (ii) The dual can be, of course, dually obtained.

12. For every $a, b, c \in N$, $a + c \neq b + c$ and $ac \neq bc$ iff $a \neq b$.

PROOF:

Since $a \neq b$ implies either $a < b$ or $a > b$, and since the duals of Prob. 11 hold simultaneously and exclusively to each other, this theorem is merely a restatement of Prob. 11, and as such is valid, of course.

Note. $a + c = b + c$ and $ac = bc$ iff $a = b$, since this is merely a contrapositive form of the theorem just proved. Note, also, that this form is the immediate result of the uniqueness of addition and multiplication in N (cf. Supplementary Problems 4.3, 4.4).

13. For every element $a \in N$, $1 \leq a$.

PROOF:

By N2, $a \neq 1$ implies $a = b' = b + 1 > 1$; and by N1-2, $a < 1$ is always false. Hence $a \geq 1$, i.e. $1 \leq a$.

14. If $a < b$ and $c < d$ for every $a, b, c, d \in N$, then $a + c < b + d$ and $ac < bd$. The dual holds, and likewise, $a + c = b + d$ and $ac = bd$ if $a = b$ and $c = d$.

PROOF:

By Prob. 11, $a < b$ implies $a + c < b + c$; likewise $c < d$ implies $c + b < d + b$. Hence $a + c < b + c = c + b < d + b$, i.e. $a + c < b + d$.

Similarly, $a + c > b + d$ and $ac > bd$ if $a > b$ and $c > d$.

The duals, which hold simultaneously and exclusively, imply $a \neq b$ and $c \neq d$ if $a + c \neq b + d$ and $ac \neq bd$. This, by MTh.1.1.1.12, completes the theorem.

15. There exists some $c \in N$ such that $a < bc$ for every $a, b \in N$.

PROOF:

The theorem holds for any $c \in N$ such that $c > a$, since, by Prob. 13, $b \geq 1$ and, by Prob. 11, 14, $bc > a \cdot 1 = a$, completing the proof.

Note. This is the familiar *Archimedean* order in N ; cf. Th.5.1.1.5 and Th.5.1.2.12 for the same order for rational numbers and real numbers.

16. There exists no $b \in N$ such that $a < b < a + 1$ for any $a \in N$.

PROOF:

If $a < b$, then $b = a + m$ for some $m \in N$. But, by Prob. 13, $1 \leq m$ and, by Prob. 11, $a + 1 \leq a + m$, i.e. $a + 1 \leq b$, while N2 contradicts $a + 1 > b$, which implies $b \leq a$ against the hypothesis. This completes the proof.

17. Prove Th. 4.1.2.3.4.

PROOF:

The complex D^+ (cf. Df. 4.1.2.2.5) of \bar{D} , which consists of positive elements of \bar{D} , contains the multiplicative identity e , which may be replaced by 1, and satisfies N1-2.

Let \bar{N} be the (unique) intersection of all subsets S_i , $i=1,2,\dots,n$, of D^+ , each of which satisfies N1-2, i.e. $a \in \bar{N}$ iff $a \in S_i$. Then \bar{N} satisfies *a fortiori* N1-2 and also N3, since $a+1 = b+1$ for every $a, b \in \bar{N}$ implies $a = b$ and conversely.

Let M be a complex of \bar{N} , satisfying N1-2; i.e. $M \subseteq \bar{N}$. Then M is one of S_i , which implies $\bar{N} \subseteq M$. Hence $\bar{N} = M$, which proves that \bar{N} satisfies N4, completing the proof.

18. Equality of elements in J is an equivalence relation.**PROOF:**

- (i) $(a, b) = (a, b)$ since $a - b = a - b$ or what is the same: $a + b = b + a$.
- (ii) $(a, b) = (c, d)$ implies $(c, d) = (a, b)$, since $a - b = c - d$ (or $a + d = b + c$) implies $c - d = a - b$ (or $b + c = a + d$).
- (iii) $(a, b) = (c, d)$ and $(c, d) = (e, f)$ imply $(a, b) = (e, f)$, since $a - b = c - d$ and $c - d = e - f$ imply $a - b = e - f$; or what is the same: $a + d = b + c$ and $c + f = d + e$ imply $a + d + c + f = b + c + d + e$, which in turn implies $a + f = b + e$, i.e. $(a, b) = (e, f)$, completing the proof.

19. For every $a, b \in J$ and $x \in N$, $(a+x, b+x) = (a, b)$.**PROOF:**

Since $(a, b) = (a, b)$ or $a + b = a + b$ in J , by Prob. 18,i above, it follows, from Prob. 3, that $a + b + x = a + b + x$ for any $x \in N$, which by Prob. 1-2 implies $(a+x) + b = (b+x) + a$. Hence, by Df. 4.1.2.3.5, $(a+x, b+x) = (a, b)$, completing the proof.

20. The binary operations in J are well-defined.**PROOF:**

Let $(a, b), (a', b'), (c, d), (c', d') \in J$ and $(a, b) = (a', b')$, $(c, d) = (c', d')$. Then, by applying Df. 4.1.2.3.6 and Prob. 19 repeatedly,

- (i) $(a', b') + (c', d') = (a' + c', b' + d') = (a + c + a' + c', a + c + b' + d')$
 $= (a + c + a' + c', (a + b') + (c + d')) = (a + c + a' + c', (a' + b) + (c' + d))$
 $= (a + c + (a' + c'), b + d + (a' + c')) = (a + c, b + d) = (a, b) + (c, d)$
 proving that $(a, b) = (a', b')$ and $(c, d) = (c', d')$ imply $(a, b) + (c, d) = (a', b') + (c', d')$;
- (ii) $(a', b')(c', d') = (a'c' + b'd', a'd' + b'c') = (ac' + bd' + a'c' + b'd', ac' + bd' + a'd' + b'c')$
 $= (ac' + bd' + a'c' + b'd', c'(a + b') + d'(b + a')) = (ac' + bd' + a'c' + b'd', c'(a' + b) + d'(b' + a))$
 $= (ac' + bd' + a'c' + b'd', c'a + c'b' + d'b + d'a') = (ac' + bd', c'b + d'a)$
 $= (ac + bd + ac' + bd', ac + bd + c'b + d'a) = (ac + bd + ac' + bd', a(c + d') + b(d + c'))$
 $= (ac + bd + ac' + bd', a(c' + d) + b(d' + c)) = (ac + bd + ac' + bd', ac' + ad + bd' + bc)$
 $= (ac + bd, ad + bc) = (a, b)(c, d)$

proving that $(a, b) = (a', b')$ and $(c, d) = (c', d')$ imply $(a, b)(c, d) = (a', b')(c', d')$.

Hence the binary operations in J are well-defined.

21. Addition in J is associative and commutative.**PROOF:**

- (i) Let $(a, b), (c, d), (e, f) \in J$; then, by Df. 4.1.2.3.4-5,

$$\begin{aligned} (a, b) + ((c, d) + (e, f)) &= (a, b) + (c + f, d + e) = (a + d + e, b + c + f) \\ &= (a + d, b + c) + (e, f) = ((a, b) + (c, d)) + (e, f) \end{aligned}$$

establishing additive associativity in J .

- (ii) Let $(a, b), (c, d) \in J$; then, by Df. 4.1.2.3.4-5,

$$(a, b) + (c, d) = (a + d, b + c) = (d + a, c + b) = (c, d) + (a, b)$$

verifying additive commutativity in J .

22. Multiplication in J is associative, commutative, and distributive under addition.

PROOF:

Let $(a,b),(c,d),(e,f) \in J$; then, by Th. 4.1.2.3.4-5,

$$\begin{aligned} \text{(i)} \quad (a,b)((c,d)(e,f)) &= (a,b)(ce+df, cf+de) = (a(ce+df)+b(cf+de), a(cf+de)+b(ce+df)) \\ &= ((ac+bd)e+(ad+bc)f, (ac+bd)f+(ad+bc)e) \\ &= (ac+bd, ad+bc)(e,f) = ((a,b)(c,d))(e,f) \end{aligned}$$

proving multiplicative associativity in J ;

$$\text{(ii)} \quad (a,b)(c,d) = (ac+bd, ad+bc) + (ca+db, da+cb) = (c,d)(a,b), \text{ proving multiplicative commutativity in } J;$$

$$\begin{aligned} \text{(iii)} \quad (a,b)((c,d)+(e,f)) &= (a,b)(c+e, d+f) = (a(c+e)+b(d+f), a(d+f)+b(c+e)) \\ &= ((ac+bd)+(ad+bf), (ad+bc)+(af+be)) = (ac+bd, ad+bc) + (ae+bf, af+be) \\ &= (a,b)(c,d) + (a,b)(e,f) \end{aligned}$$

and likewise $((a,b)+(c,d))(e,f) = (a,b)(e,f) + (c,d)(e,f)$, proving distributivity in J .

23. Prove Th. 4.1.2.3.8.

PROOF:

It follows from Df. 4.1.2.3.5 that every element of J^+ is of the form $(a+x, a)$ for every $a, x \in N$. If $(a+x, a), (b+y, b) \in J^+$, where $a, b, x, y \in N$, then, by Th. 4.1.2.3.6,

$$\begin{aligned} \text{(i)} \quad (a+x, a) + (b+y, b) &= ((a+b)+x+y, (a+b)) \leftrightarrow x+y, \\ \text{(ii)} \quad (a+x, a) \cdot (b+y, b) &= ((ay+2ab+bx)+xy, (ay+2ab+bx)) \leftrightarrow xy, \end{aligned}$$

and furthermore the following trichotomy preserves order:

$$\text{(iii)} \quad (a+x, a) \cong (b+y, b) \leftrightarrow x \cong y$$

24. There is no integer between 0 and 1.

PROOF:

Suppose $0 < x < 1$ for some $x \in I'$, where I' is a complex of the set I^+ of all positive integers. Then, by the well-ordering principle, there must exist a least element y , $y \in I'$, such that $0 < y < 1$, which implies $0 < y^2 < y$, where $y^2 \in I'$, which is manifestly a contradiction, completing the proof.

25. $x \mid 1$, where $x \in I$, iff $x = \pm 1$.

PROOF:

- (i) If $x = \pm 1$, then evidently $x \mid 1$.
- (ii) Conversely, if $ab = 1$, where $a, b \in I^+$, $a \neq 0$, $b \neq 0$, implies $a = \pm 1$ and $b = \pm 1$, then $x \mid 1$ must imply $x = \pm 1$. Now, by Th. 4.1.3.3.13, $ab = 1$ implies $|ab| = |a||b| = 1$, which in turn implies, by Prob. 16 and Df. 4.1.2.2.5, $|a| \cong 1$ and $|b| \cong 1$. Hence it must be the case that $|a| = |b| = 1$, i.e. $a = \pm 1$ and $b = \pm 1$, completing the proof.

26. Prove Th. 4.1.2.3.11.

PROOF:

- (i) Reflexivity: $a \mid a$ is always true, since $a = a \cdot 1$.
- (ii) Transitivity: $a \mid b$ and $b \mid c$ imply $a \mid c$, since, by Df. 4.1.2.2.14, there exist d_1 and d_2 , where $d_1, d_2 \in I$, such that $b = ad_1$ and $c = bd_2$, which imply $c = a(d_1d_2)$, where $d_1d_2 \in I$ by D6, i.e. $a \mid c$.
- (iii) Divisibility is not always symmetric, since $a \mid b$ and $b \mid a$ iff $a = \pm b$ (or what is the same, $b = \pm a$). For, if $a \mid b$ and $b \mid a$, then $a = bd_1$ and $b = ad_2$, as in (ii) above, which imply $a = bd_2d_1$, which in turn implies, by D11, $1 = d_2d_1$ for $a \neq 0$.

Hence, by Prob. 17, $d_1 = \pm 1$, which concludes that $a = \pm b$. (If $a = 0$, then $b = 0$, in which case it is trivially true that $a = \pm b$.) And, conversely, it is evident that $a \mid b$ and $b \mid a$ if $a = \pm b$.

27. An element u in an integral domain D is a unit in D iff its multiplicative inverse u^{-1} is also in D .

PROOF:

- (i) If u is a unit, then by definition $u|e$ and $au=e$ for some $a \in D$, proving that $a = u^{-1} \in D$.
- (ii) If $u^{-1} \in D$, then $uu^{-1}=e$, by D10, which implies $u|e$. Moreover, since $be=b$ for every $b \in D$, it follows that $ue=u$, i.e. $e|u$, completing the proof that $u|e$ and $e|u$ iff $u^{-1} \in D$.

28. For some elements $a, b \in D$, $a|b$ and $b|a$ iff a (or b) is a unit times b (or a).

PROOF:

- (i) If $a|b$ and $b|a$, then there exist two elements $c, d \in D$ such that $a=bc$ and $b=ad$, which together imply $a=adc$. On the other hand, by D10, $a=ae$. Hence, by D11, $e=dc$ which implies, by definition, c and d are units, proving that a (or b) is a unit times b (or a).
- (ii) If, say, $a=bu$, where u is a unit in D such that $uu'=e$, then $b=be=b(uu')=(bu)u'=au'$, which implies $a|b$. Similarly, $b=au$ implies $b|a$, completing the proof.

29. Prove Th. 4.1.2.2.13.

PROOF:

Let N be the set of all non-negative integers of the form $a-bq$; then $a \in N$, since $a = a - 0 \cdot b > 0$, which implies that N contains at least one positive integer. Also, by the well-ordering principle, N contains a least positive integer, say $r = a - bq$ for some non-negative integer q , where the case is exhausted by a trichotomy: $r > b$ or $r = b$ or $r < b$.

If $r > b$, then $0 < r - b = (a - bq) - b = a - b(q+1)$, which implies $(r-b) \in N$, while obviously $r-b < r$, contradicting the assumption that r is the smallest positive integer in N . Hence it must be the case that either $r=b$ or $r < b$.

If $r=b$, then $a = bq + b = b(q+1) + 0$, verifying the theorem.

If $r < b$, then $a = bq + r$ is already complete, immediately establishing the theorem.

(Second proof, by induction.

- (i) For $a=1$, the theorem evidently holds, either with $b=1, q=1, r=0$ or with $b>1, q=0, r=1$.
- (ii) Assume that $a = bq + r$, $0 \leq r < b$; then $a+1 = bq + r + 1$, where manifestly either $0 \leq r+1 < b$, in which case the theorem at once holds for $a+1$, or at most $r+1 = b$ (since $0 \leq r < b$), in which case $a+1 = bq + b = b(q+1) + 0$, again verifying the theorem for $a+1$.
- (iii) Hence the theorem holds in general.)

Furthermore, the quotient q and the remainder r are uniquely determined, since $a = bq + r = bq' + r'$, where $0 \leq r, r' < b$, implies $r - r' = b(q' - q)$, which is smaller than b , yet a multiple of b , yielding a self-contradiction. Hence it must be the case that $r - r' = 0$, i.e. $r = r'$, which implies $bq = bq'$ and, by D11, $q = q'$, establishing the uniqueness of q and r .

30. Any complex C of integers closed under addition and subtraction consists of either zero alone or all integral multiples of a least positive integer in C .

PROOF:

- (i) The case of C with zero alone is trivial.
- (ii) If $a \in C, a \neq 0$, then by hypothesis $(a-a) \in C$, i.e. $0 \in C$, and consequently also $(0-a) \in C$, i.e. $-a \in C$.

Hence there exists at least one positive element: $|a| = \pm a \in C$, which by the well-ordering principle implies a least positive element $b \in C$.

Furthermore, by induction, $kb \in C$ ($k \in I$) implies $(k+1)b \in C$ since $(k+1)b = kb + b$ and by hypothesis $kb + b \in C$, where $kb \in C$ and $b \in C$. Hence all integral multiples of b are in C .

(It can be further proved, on the strength of Prob. 31 below, that C does not contain any element other than the integral multiples of b . For the Division Algorithm of Prob. 29 concludes that any element $a \in C$ implies $a - bq = r \in C$, where $0 \leq r < b$ and b , as above, may be considered the smallest positive element of C . Hence $r=0$, which implies $a=bq$, completing the proof that the integral multiples of b exhaust C .)

31. Prove the *Euclidean Algorithm*: any two nonzero integers a and b have a positive g.c.d.
PROOF:

By Df. 4.1.2.2.5, $a > b$ or $a = b$ or $a < b$. If $a = b$, then the g.c.d. of a and b is manifestly a or b itself. Hence let $a > b$ (since the case of $a < b$ can be similarly treated); then, by Prob. 21, $a = bq + r$, $0 \leq r < b$.

- (i) If $r = 0$, then b is evidently the g.c.d. of a and b .
(ii) If $r \neq 0$, then let $d = (a, b)$ and $d' = (b, r)$, where $d \mid r$ and $d = (b, r)$, since $d \mid a$, $d \mid b$, and $d \mid (a - bq)$. Hence $d \mid d'$. Likewise, $d' \mid b$, $d' \mid r$, $d' \mid (bq + r)$, and consequently, $d' \mid a$ and $d' = (a, b)$, i.e. $d' \mid d$. Hence $d' = d$, which implies that the g.c.d. of a and b is also the g.c.d. of b and r .

Apply, then, the division algorithm to b and r , producing $b = rq_1 + r_1$, $0 \leq r_1 < r$, which implies that r is the g.c.d. of b and r (i.e. a and b) if $r_1 = 0$. If not, apply the algorithm repeatedly, and

$$\begin{aligned} b &= rq_1 + r_1 & 0 < r_1 < r \\ r &= r_1q_2 + r_2 & 0 < r_2 < r_1 \\ &\vdots \\ r_i &= r_{i+1}q_{i+2} + r_{i+2} & 0 < r_{i+2} < r_{i+1} \end{aligned}$$

which, as it goes on, must ultimately reduce r_{n+1} to zero, viz.,

$$\begin{aligned} r_{n-2} &= r_{n-1}q_n + r_n & 0 < r_n < r_{n-1} \\ r_{n-1} &= r_nq_{n+1} \end{aligned}$$

But, evidently, $(a, b) = (b, r) = (r, r_1) = \dots = (r_{n-1}, r_n) = r_n$. Hence r_n is the g.c.d. of a and b .

32. There exist integers x and y such that $d = (a, b) = ax + by$, where $a, b \in I$.

PROOF:

Since, by Prob. 31,

$$r = a - bq = a + b(-q)$$

and

$$r_1 = b - rq_1 = b - (a - bq)q_1 = a(-q_1) + b(1 + qq_1),$$

assume

$$r_i = ax_i + by_i, \quad i \in I^+,$$

of which the cases for $i=0$ and $i=1$ have already been verified as directly above. Then, from Prob. 31 and by induction,

$$\begin{aligned} r_j &= r_{j-2} - r_{j-1}q_j = ax_{j-2} + by_{j-2} - (ax_{j-1} + by_{j-1})q_j \\ &= a(x_{j-2} - x_{j-1}q_j) + b(y_{j-2} - y_{j-1}q_j) \end{aligned}$$

completing the proof.

Second Proof. The set L of all linear combinations of the form $ax + by$ is closed under addition and subtraction, since

$$(ax_1 + by_1) \pm (ax_2 + by_2) = a(x_1 \pm x_2) + b(y_1 \pm y_2)$$

Hence, by Prob. 30, L consists of all multiples of some minimal positive number $d = ax + by$, where, evidently, $c \mid a$ and $c \mid b$ imply $c \mid d$ for any $c \in I^+$. But $a, b \in L$, since $a = a \cdot 1 + b \cdot 0$ and $b = a \cdot 0 + b \cdot 1$, which implies that a and b are also the multiples of the minimal number $d \in L$, i.e. $d \mid a$ and $d \mid b$. Hence $d = (a, b) = ax + by$.

33. If p is a prime, then $p \mid ab$ implies $p \mid a$ or $p \mid b$.

PROOF:

By definition, $x \mid p$ iff $x = \pm 1$ or $x = \pm p$. Hence, if $p \nmid a$, then ± 1 are the only common divisors of p and a and consequently $1 = (a, p)$, which implies, by Prob. 32, there exist two integers x and y such that

$$1 = ax + py$$

i.e.

$$b = bax + bpy$$

where, of course, $p \mid bax$ and $p \mid bpy$. Hence $p \mid b$.

Likewise, $p \mid a$, completing the proof.

34. Prove (i) $(a, b) = 1$ and $b \mid ac$ imply $b \mid c$, and (ii) $(a, b) = 1$, $a \mid c$, and $b \mid c$ imply $ab \mid c$.

PROOF:

- (i) Since, by hypothesis and Prob. 32, $1 = ax + by$, it follows that $c = cax + cby$, where $b \mid ca$ and, of course, $b \mid cb$. Hence $b \mid c$.
- (ii) Since $a \mid c$, let $c = ad$; then, by hypothesis, $b \mid ad$, and, by (i) above, $b \mid d$, i.e. $d = bd'$, which implies $c = ad = a(bd')$. Hence $ab \mid c$.

35. If a prime p divides a product of integers $a_1 a_2 \dots a_n$, then p divides one of the factors.

PROOF:

Suppose $p \nmid a_1$; then a_1 is prime to p and, by hypothesis, $p \mid a_2 a_3 \dots a_n$. Suppose $p \mid a_2$; then, likewise, $p \mid a_3 a_4 \dots a_n$. Repeating the process, it follows that, at least, $p \mid a_n$ if p does not divide any of a_1, a_2, \dots, a_{n-1} .

36. Prove Th. 4.1.2.3.17.

PROOF:

If n is not already a prime, then, by Df. 4.1.2.3.10, there must exist two positive integers a and b such that $n = ab$, where a and b may be similarly factored if they are not primes themselves. Repeat this process until factoring is no longer feasible; then n , which is a finite number, is factored to primes alone in a finite number of steps, i.e.,

$$n = \prod_i p_i, \quad i = 1, 2, \dots, r \quad (1)$$

where p_i is a positive prime.

Suppose that the factoring is not unique, producing another expression, viz.,

$$n = \prod_j q_j, \quad j = 1, 2, \dots, s \quad (2)$$

Then, by (1) and (2),

$$p_1 p_2 \dots p_r = q_1 q_2 \dots q_s$$

which implies that any of p_i , say p_1 , divides $\prod q_j$ and, by Prob. 35, p_1 divides some q_j , say q_1 (by a rearrangement of the order of q_j if necessary). Hence $p_1 = q_1$, since q_1 is also a prime.

It follows, then, that since $p_i \neq 0$,

$$p_2 p_3 \dots p_r = q_2 q_3 \dots q_s$$

where it is deducible, as above, that $p_2 = q_2$. This process is repeated until $p_r = q_s$ is verified, where $r = s$ is also deduced, completing the proof.

Note. (1) should be more properly written as

$$p_1^{a_1} p_2^{a_2} \dots p_n^{a_n}$$

where the p 's are so arranged that $1 < p_1 < p_2 < \dots < p_n$, and the a 's are positive integers, since the occurrence of equal primes should not be excluded; e.g. $360 = 2^3 3^2 5$.

37. Prove Th. 4.1.2.3.19.

PROOF:

- (i) By Df. 4.1.2.3.18, $a \equiv b \pmod{m}$ iff $a - b = km$, $k \in I$, i.e. $m \mid (a - b)$ or $m \mid ((a + x) - (b + x))$, since $a - b = a + x - b - x$ for any $x \in I$. Hence $a \equiv b \pmod{m}$ implies $a + x \equiv b + x \pmod{m}$.
- (ii) Likewise, $m \mid (ax - bx)$, $x \in I$, if $m \mid (a - b)$. Hence $a \equiv b \pmod{m}$ implies $ax \equiv bx \pmod{m}$.

38. If $(c, m) = 1$, then $ca \equiv cb \pmod{m}$ implies $a \equiv b \pmod{m}$.

PROOF:

Since, by hypothesis, $m \mid (ca - cb)$, i.e. $m \mid c(a - b)$, and also $(c, m) = 1$, it follows at once from Prob. 34(i) that $m \mid (a - b)$, i.e. $a \equiv b \pmod{m}$.

(Cf. Th. 3.2.6.9, and note that the cancellation law does not generally hold for congruences.)

39. If $d = (c, m)$, then $ac \equiv bc \pmod{m}$ implies $a \equiv b \pmod{m/d}$.

PROOF:

Since $d = (c, m)$, it follows at once that $c = c'd$ and $m = m'd$ where $(c', m') = 1$. But, by hypothesis, $m \mid (ac - bc)$, i.e. $m'd \mid c'd(a - b)$, which implies $m' \mid (a - b)$. Hence $a \equiv b \pmod{m'}$, i.e. $a \equiv b \pmod{m/d}$.

40. If $(a, m) = 1$, then $ax \equiv b \pmod{m}$ has a unique solution modulo m .

PROOF:

By hypothesis and Prob. 32, $1 = ap + mq$, which implies $b = bap + bmq$, i.e. $a(bp) - b = (-bq)m$, which in turn implies $abp \equiv b \pmod{m}$. Hence bp is a solution of $ax \equiv b \pmod{m}$.

Suppose that y is also a solution of the given congruence; then, by Th. 3.2.6.8, $ay - abp \equiv b - b \equiv 0 \pmod{m}$. Hence, by Prob. 39, $y \equiv bp \equiv x \pmod{m}$, completing the proof.

41. Prove Th. 4.1.2.3.20.

PROOF:

(i) Let $d = (a, m)$, i.e. $a = a'd$ and $m = m'd$ where $(a', m') = 1$; then, if y is a solution of the congruence, $ay - b = km$, by Df. 4.1.2.3.18. Hence $b = ay - km = a'dy - km'd = d(a'y - km')$, which implies $d \mid b$.

Conversely, if $d \mid b$, i.e. $b = b'd$, then, by Prob. 40, $a'x \equiv b' \pmod{m'}$ has a unique solution modulo m' , since $(a', m') = 1$.

Hence $ax \equiv b \pmod{m}$ has an integral solution x iff $(a, m) \mid b$.

(ii) If x_i is such a solution as prescribed as in (i), then it represents a class of solutions $a'x \equiv b' \pmod{m'}$, x_i representing

$$x_1, \quad x_2 = x_1 + m', \quad \dots, \quad x_d = x_1 + (d-1)m'$$

since, for every k , $x_i + km'$ is also a solution, because

$$a(x_i + km') = ax_i + a'dkm' = a'km, \quad \text{i.e. } ax_i \equiv b \pmod{m}$$

Hence any of x_i , $i = 1, 2, \dots, d$ is a solution.

Moreover, each of x_i is unique, since the assumption $x_i \equiv x_j \pmod{m}$ for $i \neq j$, say $i > j$, implies

$$x_i = x_1 + (i-1)m' - x_1 + (j-1)m' = x_j \pmod{m}$$

which in turn implies $(i-j)m' \equiv 0 \pmod{m}$, i.e. $m \mid (i-j)m'$, yielding a contradiction, since $0 < i-j < d$ such that $0 < (i-j)m' < dm' = m$.

Hence no two of the d solutions are congruent modulo m .

Finally, the set of x_i evidently exhausts the solutions, since every solution x_k of the congruence is congruent to x_i modulo m' (since $ax_k \equiv b$ and $ax_i \equiv b \pmod{m}$, i.e. $ax_k \equiv ax_i \pmod{m}$), must imply, by Prob. 39, $x_k \equiv x_i \pmod{m'}$.

42. Prove Th. 4.1.2.3.21.

PROOF:

The first congruence has a solution a_1 and, in general, $a_1 + bm_1$ for any $b \in I$, which must satisfy the second congruence, viz. $a_1 + bm_1 \equiv a_2 \pmod{m_2}$, i.e. $bm_1 \equiv a_2 - a_1 \pmod{m_2}$, which is actually solvable for b by Th. 4.1.2.3.20, since $(m_1, m_2) = 1$.

Conversely, any two solutions x and x' of the given simultaneous congruences imply $x - x' \equiv 0$ for both $\pmod{m_1}$ and $\pmod{m_2}$, which in turn implies, since $(m_1, m_2) = 1$, that $m_1 m_2 \mid (x - x')$, i.e. $x \equiv x' \pmod{m_1 m_2}$, completing the proof.

43. The simultaneous congruences of the form $x \equiv a_i \pmod{m_i}$, $i = 1, 2, \dots, n$, have a common solution iff $a_i \equiv a_j \pmod{(m_i, m_j)}$, $j = 1, 2, \dots, n$, where $(m_i, m_j) = 1$, and the solution is unique, modulo $\prod m_i$.

PROOF:

Let $M = \prod m_i$ and $M_i = M/m_i$; then, by hypothesis, $(m_i, M_i) = 1$, which implies that there exist integers $b_i \in I$ such that $M_i b_i \equiv 1 \pmod{m_i}$. If $c \equiv \sum a_i M_i b_i$, then $c \equiv a_1 M_1 b_1 \equiv a_1 \pmod{m_1}$, since M_2, M_3, \dots, M_n are all multiples of m_1 . Likewise, in general,

$$x \equiv a_i M_i b_i \equiv a_i \pmod{m_i}$$

is a solution of the given simultaneous congruences.

Conversely, if $x = d$, then it follows at once from the above result that

$$d \equiv c \pmod{m_i}, \quad \text{i.e. } d \equiv c \pmod{\prod m_i}$$

which implies that $d - c$ is a common multiple of m_i and that the least common multiple is divisible by $\prod m_i$. Hence $d \equiv c \pmod{\prod m_i}$, completing the proof.

4.1.2.4 FIELDS IN GENERAL

Df. 4.1.2.4.1 A ring is a *field*, denoted by F , if it forms an Abelian group under both addition and multiplication and satisfies the distributive laws under addition:

- (i) $a(b + c) = ab + ac$, (ii) $(a + b)c = ac + bc$, for every $a, b, c \in F$.

Stated in detail, a set F is a field if it satisfies the following eleven axioms:

F1-8 \equiv **R1-8**

F9: Multiplicative identity, i.e. unity e (or 1), which is defined by $ea = ae = a$ (or $1 \cdot a = a \cdot 1 = a$) for every $a \in F$.

F10: Multiplicative inverse, a^{-1} , which uniquely exists for every $a \in F$, where $a \neq 0$, such that $aa^{-1} = a^{-1}a = e$.

F11: Multiplicative commutativity, $ab = ba$ for every $a, b \in F$.

A field, then, is an integral domain with the additional property of **F10**, which is indeed more restricting than **D11** (cf. Df. 4.1.2.2.1).

Example:

The set R of rational numbers, the set \bar{R} of real numbers, and the set C of complex numbers are all fields; on the other hand, the integers fail to form a field, evidently unable to satisfy **F10**, while the residue class modulo m forms a field iff m is a prime (cf. Prob. 3-4 below). Thus $I/\{2\}, I/\{3\}, I/\{5\}$, etc. are all fields, while $I/\{4\}, I/\{6\}$, etc. are not even integral domains (cf. Th. 4.1.2.2.5), let alone fields.

Df. 4.1.2.4.2 A *subfield* S of a field F is a complex of F which is itself a field.

Example: R in the preceding example is a subfield of \bar{R} , which in turn is a subfield of C .

In view of these examples, the definition of subfields may be made more articulate, viz.:

Df. 4.1.2.4.2a A complex F' of a field F determines a subfield iff:

- (i) $a + b \in F'$ for every $a, b \in F'$, and $a \in F'$ implies $-a \in F'$; also $0 \in F'$.
(ii) $a \cdot b \in F'$ for every $a, b \in F'$, $a \neq 0, b \neq 0$, and $a \in F'$ implies $a^{-1} \in F'$; also $1 \in F'$.

As can be readily examined in terms of (i) and (ii), the meet of any number of subfields of F is again a subfield (cf. §4.2.1, Prob. 4), which is evidently the smallest of the subfields and is called the *prime* (or *minimal*) subfield. This name originates from the following definition:

Df. 4.1.2.4.2b A field is called a *prime field* if it contains no proper subfield, i.e. no subfield other than itself.

Example:

The set R of rational numbers is the prime subfield of F if F is of characteristic zero, and the residue class $I/\{p\}$ modulo a prime p is also the prime subfield of F if F is of characteristic p (cf. Th. 4.2.1.4); i.e. F has two kinds of prime subfields, since the characteristic of F is either zero or a prime (cf. Th. 4.1.2.4.3 below and also §5.1.1, Prob. 2-5).

Th. 4.1.2.4.3 A field is necessarily an integral domain. (Cf. Prob. 2 and also Th. 4.1.2.4.4 below.)

Note. The characteristic of a field, then, is *a fortiori* either zero or a prime (cf. Th. 4.1.2.2.5).

Th. 4.1.2.4.4 Any finite integral domain is a field. (Cf. Prob. 6.)

A field with a finite number of elements is called a *finite field* or a *Galois field* (cf. Prob. 7), the study of which, however, goes beyond the scope of the present text.

Th. 4.1.2.4.5 Division, except by zero, is possible in F , producing a unique result, viz. a solution of the form b/a for $ax = b$, where $a, b, x \in F$. (Cf. Prob. 8.)

The solution, defined by Th. 4.1.2.4.5, is called the *quotient* of the division, which has the following specific properties:

Th. 4.1.2.4.6 If $a, b, c, d \in D$ and $b \neq 0, d \neq 0$, then

$$(i) \quad (a/b) = (c/d) \text{ iff } ad = bc$$

$$(ii) \quad (a/b) \pm (c/d) = (ad \pm bc)/(bd)$$

$$(iii) \quad (a/b)(c/d) = (ac/bd). \quad (\text{Cf. Prob. 9.})$$

It should be emphasized here that the form of *fractions*: $a/b, c/d$, etc. is not necessarily inherent in the concept of the quotient, which indeed can be equally well expressed in terms of pairs: $(a, b), (c, d)$, etc. or products: ab^{-1}, cd^{-1} , etc.

Example:

$$(ad \pm bc)/(bd) = (ad \pm bc, bd) \text{ or } (ad \pm bc)(bd)^{-1}, \text{ etc.}$$

Th. 4.1.2.4.7 A set Q , whose elements are of the form defined by Th. 4.1.2.3.6.i, forms a field. (Cf. Prob. 11.)

Df. 4.1.2.4.7a The field Q of Th. 4.1.2.4.7 is called the *quotient field* of an integral domain D .

Th. 4.1.2.3.8 The quotient field Q of an integral domain D contains a complex Q' , which is isomorphic to D . (Cf. Prob. 13.)

It follows directly from this theorem that, if the integral domain D is already a subfield of a field F , the quotient field Q of D is then isomorphic to a subfield of F ; e.g. Q is isomorphic to the field of rational numbers, a subfield of F , if F is the field of real numbers and D the integral domain of integers, and conversely, as is in the following theorem.

Th. 4.1.2.4.9 The set R of all rational numbers, which consists of number couples (a, b) of integers a and $b \neq 0$, is an integral domain. (Cf. Prob. 14.)

It must be emphasized again that the form of number couples in this context is interchangeable with the form of quotients or products (cf. Th. 4.1.2.4.6). Because it is not to be mistaken for “(rational) integers” (cf. Df. 4.1.2.3.5) in the form of ordered pairs, however, the quotient form is preferred in the present text.

The set R as such, which is isomorphic to the quotient field Q , is also called the *rational number field* (cf. Df. 5.1.1.1); in the same sense, the set \bar{R} of real numbers and the set C of complex numbers are called the *real number field* and the *complex number field*, respectively. Some other fields are found by the following definitions.

Df. 4.1.2.4.10 A *quadratic field*, satisfying F1-11, consists of elements of the form $a + b\bar{r}$, where $a, b \in R$ and \bar{r} is an irrational root of a quadratic equation

$$px^2 + qx + r = 0,$$

for $p, q, r \in R$ and $p \neq 0$. The field as such is denoted by $R[1, \bar{r}]$, or simply $R[\bar{r}]$, representing the set of linear polynomials in \bar{r} .

Example:

$R[\sqrt{2}]$ designates the set of all numbers of the form $a + b\sqrt{2}$, which does form a field, as can be readily verified (cf. Prob. 1, and also §4.1.1, Prob. 10,v).

Df. 4.1.2.4.11 A *cubic field*, satisfying **F1-11**, consists of elements of the form $a + b\bar{r} + c\bar{r}^2$, where $a, b, c \in R$ and \bar{r} is a root of an irreducible (i.e. not factorable) cubic equation

$$x^3 + ax^2 + bx + c = 0$$

for $a, b, c \in R$. The cubic field is denoted by $R[1, \bar{r}, \bar{r}^2]$, or simply $R[\bar{r}^2]$, since it causes no confusion in generality.

Example:

$R[\sqrt[3]{5}]$ designates the set of numbers of the form $a + b\sqrt[3]{5} + c\sqrt[3]{25}$, which satisfies all of **F1-11**, as can be readily verified.

Quadratic and cubic fields will be observed again when they reappear as *extensions* (cf. Df. 5.3.1.5) of a field.

Since a field is *a fortiori* an integral domain, it may be ordered, as in the following definition:

Df. 4.1.2.4.12 An *ordered field* \bar{F} is a field which contains a subset F^+ of “positive” elements satisfying the additive and multiplicative closure and trichotomy of Df. 4.1.2.2.5.

Example:

The set R of rational numbers and the set \bar{R} of real numbers are both ordered (cf. Th. 5.1.1.4 and Th. 5.1.2.11-12).

The Archimedean order in the set N of natural numbers (cf. §4.1.2.3, Prob. 15), too, may be generalized to \bar{F} , viz.:

Df. 4.1.2.4.13 An ordered field \bar{F} is said to be *Archimedean* (or *Archimedean ordered*) if there exists a positive integer n such that $ne > a$ for a multiplicative unity $e \in \bar{F}$ and every $a \in \bar{F}$.

Example:

R and \bar{R} , as well as N and I , are Archimedean ordered (cf. Th. 5.1.1.6 and Th. 5.1.2.12).

Solved Problems

1. The set C of all elements of the form $a + br$, where a and b are rational numbers and r is an imaginary cube root of unity, i.e. $(-1 + 3i)/2$, is a subfield of the field of all complex numbers.

PROOF:

F1: $(a + br) + (c + dr) = (a + c) + (b + d)r \in C$, by which **F2** and **F5** are readily verified.

F3: Additive identity. $(0 + 0r) \in C \equiv 0 \in F$.

F4: Additive inverse. $-a - br$.

F6: $(a + br) \cdot (c + dr) = ac + (bc + ad)r + bdr^2 = ac + (bc + ad)r + bd(-r - 1) = (ac - bd) + (ad + bc - bd)r \in C$, from which **F7-8** and **F11** are readily obtained.

F9: Multiplicative identity. $(1 + 0r) \in C \equiv 1 \in F$.

F10: Multiplicative inverse. $-(b - a + br)/(a^2 - ab + b^2)$, where $a^2 - ab + b^2 \neq 0$ ($\because a^2 - ab + b^2 = ((a^2 + b^2) + (a - b)^2)/2 > 0$ unless $a = b = 0$, which reduces the problem to a triviality).

2. Prove Th. 4.1.2.4.3.

PROOF:

Suppose $a, b \in F$ such that $ab = 0$; then $a \neq 0$ implies $a^{-1} \in F$ and $a^{-1}(ab) = (a^{-1}a)b = eb = b$. But also, by the initial assumption, $a^{-1}(ab) = a^{-1}0 = 0$. Hence $b = 0$, proving that $ab = 0$ entails $b = 0$ or $a = 0$, which is logically equivalent to **D11** (cf. §4.1.2.2, Prob. 1). This completes the proof, since other axioms are interchangeable.

3. The set E of all even integers forms a commutative ring, but not a field.**PROOF:**

Since the sum and product of two even integers is again even, E satisfies **R1,6**, which in this case entails also **R2,7,8,5**, and multiplicative commutativity.

Furthermore, since $0 \in E$ by definition, **R3** is satisfied; so is **R4**, since every element $a \in E$ implies $-a \in E$. Hence E is a commutative ring.

E is not a field, however, since it cannot satisfy, for instance, **F9**; for e such that $ea = ae = a$ for every $a \in E$ implies $1 \in E$, which is contradictory to the definition of E .

4. The ring M of integers modulo m is a field iff m is a prime.**PROOF:**

It has already been proved (cf. §4.1.1, Prob. 8-9) that the residue class modulo m , where m is any positive integer, generally forms a commutative ring.

Assume that m is not a prime, i.e. $m = m_1 m_2$ where $m_1 > 1$ and $m_2 > 1$; then, by definition of congruence, $m_1 m_2 = 0 \pmod{m}$ with $m_1 \neq 0$ and $m_2 \neq 0$, which contradicts the main property of fields, which are supposed to have no proper zero-divisors. Hence, if M is a field, m must be a prime.

Conversely, if m is a prime, the ring M consists of the disjoint complexes: $\{0\}, \{1\}, \{2\}, \dots, \{m-1\}$, \pmod{m} . Then, since a product of integers is divisible by m iff one of the factors is divisible by m , it follows at once that M cannot contain any proper zero-divisors. This in turn implies that the set of all elements of M of the form ax , where $a \neq 0 \in M$ and x runs through the nonzero elements of M , are $m-1$ in number and all distinct, exhausting the nonzero elements of M . Hence the equation $ay = b$ has a unique solution for every $b \in M$, $b \neq 0$, including a trivial solution $x = 0$ in case $b = 0$, proving that M is a field.

5. An integral domain D whose characteristic is a prime p contains a complex C isomorphic to the field F of residue class modulo p .**PROOF:**

Since the prime order of the cyclic group generated by the unity e of D implies (cf. §4.1.2.2, Prob. 4-6) that the elements of the form ke , where k belongs to the same residue class modulo p , are equal, it immediately follows that the correspondence $c = ke \in C \leftrightarrow F_k \subset F$, where F_k denotes the residue class containing the integer k , is an isomorphism. (Stated in detail, $k_1 e \leftrightarrow F_{k_1} \pmod{p}$ and $k_2 e \leftrightarrow F_{k_2} \pmod{p}$ do imply $k_1 e + k_2 e = (k_1 + k_2)e \leftrightarrow F_{k_1 + k_2} \pmod{p}$ and $(k_1 e)(k_2 e) = (k_1 k_2)e \leftrightarrow F_{k_1 k_2} \pmod{p}$, by Th. 3.2.6.8.)

6. Prove Th. 4.1.2.4.4.

PROOF:

Since D is finite, assume that d_1, d_2, \dots, d_n are the n distinct elements of D and let F be a set whose elements are $d_k d_1, d_k d_2, \dots, d_k d_n$, where $d_k \neq 0 \in D$ is fixed. Then, D being an integral domain, all elements of F are distinct ($\because d_k d_i = d_k d_j$ implies $d_i = d_j$ by **D11**). F has thus n distinct elements, and since D itself has the n distinct elements, it implies $F = D$.

Hence the unity 1 of D must exist somewhere among the elements of F as $d_k d_m = e$, $1 \leq m \leq n$, which implies d_m is the inverse of d_k . This proves F to be a field, since any specific element $d_k \neq 0$ in $D = F$ has an inverse.

7. If a field F is finite and has n elements, then n is of the form p^m , where p is the prime characteristic of F and m is any positive integer.

PROOF:

- (i) Given $e, 2e, \dots, (n+1)e$, where e is the unity of F , it is evident that these multiples of e cannot be distinct, since F has only n elements. Suppose, then, any two different integers a and b , say $a > b$, such that $ae = be$. This entails at once $(a-b)e = 0$, which in turn implies that F must be of a prime characteristic, say p (cf. Prob. 4 above).

Hence the p elements of a set F' of the form $ie, i=1,2,\dots,p-1$, are distinct and every integral multiple of e must be found in F' . If $F=F'$, this completes the proof for $m=1$.

- (ii) Suppose there exists an element c of F such that $c \neq ie, i=1,2,\dots,p-1$, and that a set F'' has p^2 elements of the form $ie + jc, i,j=1,2,\dots,p-1$. If some elements of F'' are not distinct, then, for $i',j'=1,2,\dots,p-1$,

$$ie + jc = i'e + j'c, \quad \text{i.e. } (i-i')e = (j'-j)c \quad (1)$$

assuming $j \neq j'$, say $j' > j$, and $p > j' - j > 0$. Hence, by Prob. 4 and §4.1.2.3, Prob. 32, there exist two integers a and b such that

$$a(j' - j) = 1 + bp \quad (2)$$

Multiplying (1) by a and (2) by c ,

$$a(j' - j)c = a(i - i')e \quad (3)$$

$$a(j' - j)c = c + cbp \quad (4)$$

Since $bp=0$, it follows from (3) and (4) that

$$c = a(i - i')e$$

which contradicts the assumption that $c \neq ie, i=1,2,\dots,p-1$. Hence it must be the case that $j = j'$, which implies $(i - i')e = 0$, i.e. $i = i'$, since both i and i' are evidently less than the characteristic p . This completes the proof for $m=2$.

- (iii) If $F'' \neq F$, i.e. if F contains an element d such that $d \neq ie + jc$, then all p^3 elements of the form $ie + jc + kd, i,j,k=1,2,\dots,p-1$, can be proved likewise to be distinct.

The similar procedure may be repeated, but not necessarily more than m times, m being a certain positive integer, since F is a finite field. Hence, in general, n is of the form p^m if F has n elements.

8. Prove Th. 4.1.2.4.5.

PROOF:

The equation $ax = b, a \neq 0$, in F has a solution x , since $a \neq 0$ entails a^{-1} in F , by F10, and $x = a^{-1}b$ (or what is the same: b/a). This solution is also unique, since $ax = b$ and $ay = b$ imply $ax = ay$, which in turn implies $(aa^{-1})x = (aa^{-1})y$, i.e. $x = y$, completing the proof.

9. Prove Th. 4.1.2.4.6.

PROOF:

- (i) If $ad = bc$, then

$$a/b = ab^{-1} = add^{-1}b^{-1} = bcd^{-1}b^{-1} = cd^{-1} = c/d$$

Conversely, if $a/b = c/d$, i.e. $ab^{-1} = cd^{-1}$, then

$$ad = a(bb^{-1})d = (ab^{-1})db = cd^{-1}db = bc$$

- (ii) Since $x = a/b$ and $y = c/d$ are the unique solutions of $bx = a$ and $dy = c$, it follows that $dbx = da$ and $bdy = bc$, which together imply $bd(x \pm y) = ad \pm bc$, i.e. $x \pm y = (ad \pm bc)/bd$, $bd \neq 0$ ($\because b \neq 0$ and $d \neq 0$ by hypothesis). Hence

$$(a/b) \pm (c/d) = (ad \pm bc)/bd$$

- (iii) Since, as above, $x = a/b$ and $y = c/d$ are the unique solutions of $bx = a$ and $dy = c$,

$$(bd)(xy) = (bx)(dy) = ac, \quad \text{i.e. } xy = (ac)/(bd)$$

Hence $(a/b)(c/d) = (ac)/(bd)$.

10. Prove, in the same context as above:

- (i) $(a/b) \neq 0$ implies $(a/b)(b/a) = 1$, and
- (ii) $(a/b) + (-a/b) = 0$.

PROOF:

- (i) Since, from Prob. 9, (iii) above, $(a/b)(c/d) = (ac)/(bd)$, it immediately follows that $(a/b)(b/a) = (ab)/(ba)$. But, by Th. 4.1.2.3.5, ab/ba is then the unique solution of $bax = ab$, where evidently $x = 1$. Hence $(a/b)(b/a) = 1$.
- (ii) Since, from Prob. 9, (ii) above, $(a/b) \pm (c/d) = (ad \pm bc)/bd$, it follows at once that

$$(a/b) + (-a/b) = (ab - ba)/b^2 = 0/b^2 = 0 \cdot (b^2)^{-1} = 0$$

11. Prove Th. 4.1.2.4.7.

PROOF:

F1 is already proved by Th. 4.1.2.4.6, (ii).

$$\begin{aligned} \text{F2: } (a/b) + ((c/d) + (e/f)) &= (a/b) + (cf + de)/df = (adf + bcf + bde)/bdf \\ &= (ad + bc)/bd + (e/f) = ((a/b) + (c/d)) + (e/f) \end{aligned}$$

$$\text{F3: } (0/b)$$

$$\text{F4: } (-a/b) \quad (\text{cf. Prob. 10, ii})$$

$$\text{F5: } (a/b) + (c/d) = (ad + bc)/bd = (cb + da)/db = (c/d) + (a/b)$$

$$\text{F6: } (a/b)(c/d) = (ac)/(bd) \in Q \quad (\text{cf. Th. 4.1.2.3.6, iii})$$

$$\text{F7: } (a/b)((c/d)(e/f)) = (ace)/(bdf) + ((a/b)(c/d))(e/f)$$

$$\begin{aligned} \text{F8: } (a/b)((c/d) + (e/f)) &= (a/b)((cf + de)/df) = (acf + ade)/bdf = (abcf + abde)/bbdf \\ &= (ac/bd) + (ae/bf) = (a/b)(c/d) + (a/b)(e/f) \end{aligned}$$

$$\text{F9: } (a/a) \text{ for every } a \in D \text{ (i.e. the multiplicative identity of } Q \text{ is not unique).}$$

$$\text{F10: } (a/b)^{-1} = (b/a) \quad (\text{cf. Prob. 10, i})$$

$$\text{F11: } (a/b)(c/d) = (ac)/(bd) = (c/d)(a/b)$$

12. Given $ax = b$ in Q , where $a, b \in D$ and $a \neq 0$, the solution $x = b/a$ is unique.

PROOF:

Since $ax = a(b/a) = (ac/c)(b/a) = (bac/ac) = b$, $x = b/a$ is a solution. If $y = b'/a'$ is another solution, then

$$(ac/c)(b'/a') = (bd/d), \quad cd \neq 0, \quad \text{i.e. } (acb'/ca') = (bd/d)$$

which implies, by Th. 4.1.2.4.6, (i), $acb'd = ca'bd$, which in turn implies, by D11, $ab' = a'b$. Hence, again by Th. 4.1.2.4.6, (i), $a/b = a'/b'$, completing the proof.

13. Prove Th. 4.1.2.4.8.

PROOF:

Let $a, b, c, \dots \in D$ and $(a/q), (b/q), (c/q), \dots \in Q'$, where $q \neq 0 \in D$, such that a correspondence can be set up as follows:

$$a \leftrightarrow (a/q), \quad b \leftrightarrow (b/q), \quad c \leftrightarrow (c/q), \quad \dots$$

Then Q' is isomorphic to D , since

$$(a/q) + (b/q) = ((aq + bq)/q^2) = ((a + b)/q) \leftrightarrow a + b$$

and

$$(a/q)(b/q) = (ab/q^2) \leftrightarrow (ab/q') \leftrightarrow ab$$

where evidently $q' = q^2 \neq 0$, or more generally $q' = q^n \neq 0$, $n \in I^+$, as long as $q \neq 0$, which is exactly the case in this context, completing the proof.

14. Prove Th. 4.1.2.4.9.

PROOF:

Rewriting Th. 4.1.2.4.6 in terms of couples:

- (i) $(a_1, b_1) = (a_2, b_2)$ iff $a_1 b_2 = a_2 b_1$
- (ii) $(a_1, b_1) + (a_2, b_2) = (a_1 b_2 + a_2 b_1, b_1 b_2)$
- (iii) $(a_1, b_1) \cdot (a_2, b_2) = (a_1 a_2, b_1 b_2)$

where $a_i, b_i \in I$, $b_i \neq 0$, $i = 1, 2, \dots$, and $b_1 b_2 \neq 0$, I containing no zero divisors.

Since (i) is not of formal identity (i.e. $(a_1, b_1) = (a_2, b_2)$ iff $a_1 = a_2$ and $b_1 = b_2$), it must be proved that (i) is a relation of logical equivalence, which is verified as follows:

- (i₁) $(a_1, b_1) = (a_1, b_1)$, since evidently $a_1 b_1 = a_1 b_1$.
- (i₂) $(a_1, b_1) = (a_2, b_2)$ implies $(a_2, b_2) = (a_1, b_1)$, and conversely, since $a_1 b_2 = a_2 b_1$ implies $a_2 b_1 = a_1 b_2$, and conversely.
- (i₃) $(a_1, b_1) = (a_2, b_2)$ and $(a_2, b_2) = (a_3, b_3)$ imply $(a_1, b_1) = (a_3, b_3)$, since $a_1 b_2 = a_2 b_1$ and $a_2 b_3 = a_3 b_2$ imply $a_1 b_2 b_3 = a_2 b_1 b_3 = a_3 b_1 b_2$ which, cancelling $b_2 \neq 0$, in turn implies $a_1 b_3 = a_3 b_1$.

On the strength of (i₁)-(i₃), the properties of D1-11 are proved one after another in the same way as Th. 4.1.2.4.7 above, changing only the form of quotients into that of corresponding couples.

Note. R , the rational number field, can satisfy F1-11 as well as D1-11, as is quite obvious in Th. 4.1.2.4.7. What may not be quite obvious, however, is that the proof of Th. 4.1.2.4.9 needs only the properties of I as an integral domain and is not in need of the well-ordering principle; in other words, the construction of rationals from integers can be carried out without bothering the principle.

4.1.2.5 POLYNOMIALS IN GENERAL

Df. 4.1.2.5.1 A *polynomial* in x over a commutative ring R , denoted by $p(x)$, is the following expression:

$$\sum_i a_i x^i, \quad i = 0, 1, \dots, n$$

$$\text{i.e.} \quad p(x) \equiv a_0 x^0 + a_1 x^1 + \dots + a_n x^n \quad (1)$$

where the *coefficients* a_i , not all of them zero, belong to R , while the *indeterminates*, $x_i \notin R$, are considered here commutative with every element $a \in R$, i.e. $ax^i = x^i a$. Each of $a_i x^i$ is called a *term* of $p(x)$.

Example:

$-2x + 3x^2$, i.e. $0x^0 + (-2)x^1 + 3x^2$, is a polynomial in x over I , the ring of integers, and $\sqrt{2} + 5x^4/6$, i.e. $\sqrt{2}x^0 + 0x^1 + 0x^2 + 0x^3 + 5x^4/6$, is a polynomial in x over a field F .

Note. $p(x)$ is considered unchanged, as is evident in the examples above, by the omission or insertion of the terms with zero coefficients.

Df. 4.1.2.5.2 $R[x]$ (or $R(x)$ if it is unmistakable in the context) denotes the set of all polynomials in x over R . This process of adjoining x to R in order to form $R[x]$ is called *ring adjunction* (cf. Df. 5.3.1.6 and Df. 5.3.2.1).

Example:

$$p(x), q(x), r(x), \dots \in R[x].$$

Df. 4.1.2.5.3 Two polynomials $p(x) = \sum_i a_i x^i$ and $q(x) = \sum_i b_i x^i$, $i = 0, 1, \dots, n$, are equal iff the coefficients of each superscript of x are the same, i.e. $p(x) = q(x)$ iff $a_i = b_i$, $i = 0, 1, \dots, n$.

Example:

$$2 + 7x = 2 + ax \quad \text{implies} \quad a = 7.$$

Df. 4.1.2.5.4 Given two polynomials $p(x) = \sum_i a_i x^i$, $i = 0, 1, \dots, n$, and $q(x) = \sum_j b_j x^j$, $j = 0, 1, \dots, m$, their *sum* is a polynomial $r(x) = \sum_k c_k x^k$, where $c_k = a_k + b_k$, $k = 0, 1, \dots, n$ or m , depending on $n > m$ or $n < m$, i.e.,

$$\begin{aligned} r(x) &= p(x) + q(x) \\ &= (a_0 x^0 + a_1 x^1 + \dots + a_n x^n) + (b_0 x^0 + b_1 x^1 + \dots + b_m x^m) \\ &= (a_0 + b_0)x^0 + (a_1 + b_1)x^1 + \dots + (a_m + b_m)x^m + a_{m+1}x^{m+1} + \dots + a_n x^n \end{aligned}$$

if $n > m$, and if $n < m$,

$$= (a_0 + b_0)x^0 + (a_1 + b_1)x^1 + \dots + (a_n + b_n)x^n + b_{n+1}x^{n+1} + \dots + b_m x^m$$

where a superscript of x which does not appear in $p(x)$ or $q(x)$ may be regarded as having a zero-coefficient; similarly, $n > m$ implies $b_{m+1} = \dots = b_n = 0$ and $n < m$ implies $a_{n+1} = \dots = a_m = 0$. If the latter is clearly understood, then $r(x)$ may be more simply expressed as follows:

$$r(x) = \sum_k a_k x^k + \sum_k b_k x^k = \sum_k (a_k + b_k) x^k, \quad k = 0, 1, \dots, n \text{ (or } m)$$

Example:

$$\begin{aligned} (2x^0 + 3x^1) + (0x^0 + (-4)x^1 + 5x^2) &= (2+0)x^0 + (3+(-4))x^1 + (0+5)x^2 \\ &= 2x^0 + (-1)x^1 + 5x^2 \end{aligned}$$

Df. 4.1.2.5.5 Given two polynomials $p(x)$ and $q(x)$ as above, their *product* is a polynomial $s(x)$:

$$\begin{aligned} s(x) &= p(x) \cdot q(x) \\ &= a_0 b_0 x^0 + (a_0 b_1 + a_1 b_0) x^1 + \dots + (a_0 b_k + a_1 b_{k-1} + \dots + a_k b_0) x^k + \dots + a_n b_m x^{n+m} \end{aligned}$$

where $a_i = 0$ if $i > n$, and $b_j = 0$ if $j > m$.

Example:

$$\begin{aligned} (2x^0 + 3x^1) \cdot (0x^0 + (-4)x^1 + 5x^2) &= (2 \cdot 0)x^0 + (2 \cdot (-4) + 3 \cdot 0)x^1 + (2 \cdot 5 + 3 \cdot (-4))x^2 + (3 \cdot 5)x^{1+2} \\ &= 0x^0 + (-8)x^1 + (-2)x^2 + 15x^3 \end{aligned}$$

Th. 4.1.2.5.6 The set $R[x]$, defined by Df. 4.1.2.5.2, forms a ring under the operations defined by Df. 4.1.2.5.4-5. (Cf. Prob. 1.)

Th. 4.1.2.5.7 The ring $R[x]$ of Th. 4.1.2.5.6 is commutative. (Cf. Prob. 2.)

These theorems and definitions suggest that many properties of a ring R pass over to the polynomial ring $R[x]$, and they do. Thus $R[x]$ is a ring with unity if R is with unity (cf. Prob. 3), and $R[x]$ is a ring without zero-divisors if R is without them (cf. Prob. 4). This is summarized in the following theorem. (The polynomials over a field, on the other hand, will be studied in §5.2ff.)

Th. 4.1.2.5.8 The set $R[x]$ of all polynomials over a commutative ring R with unity and without zero-divisors forms an integral domain iff R is an integral domain itself. (Cf. Prob. 5.)

This theorem directly points out a concrete method to construct various integral domains, since it gives at once that $J[x]$, $R[x]$, $R^*[x]$, and $J_p[x]$ which represent the sets of polynomials over integers, rationals, reals, and integers modulo p , a prime, respectively, are also integral domains.

Since, for instance, $f(x) = x^p - x$ and $g(x) = 0$ in $J_p(x)$ are distinct in *form*, yet the same in *function* (cf. Th. 3.2.6.13 and Supplementary Problem 4.18 for a concrete example), the following definition is stipulated at this juncture.

Df. 4.1.2.5.9 The mapping (i.e. function) $M: a \in D \rightarrow f(a)$, where $f(a)$ is called the *value* of $f(x) \in D[x]$ when x equals a , is called a *polynomial function*.

Every polynomial in $D[x]$, then, defines a polynomial function of D into D itself ($\because f(a) \in D$ if $a \in D$); in this sense, $f(x) \in D[x]$ denotes both a polynomial over D and a polynomial function of D into D . This distinction is explicit in the following theorem.

Th. 4.1.2.5.10 If D is an infinite integral domain, then distinct polynomial functions of $D[x]$ define distinct mappings of D into D . (Cf. Prob. 6.)

This is not the case, however, if D is finite (cf. Prob. 7). Th. 4.1.2.5.10 thus has a logically equivalent form: $f(x)$ and $g(x)$ over an infinite integral domain are equal if they define the same function.

Hence, if D is infinite, the polynomial ring in D is isomorphic to the ring of polynomial functions in D . Since any ring isomorphic to an integral domain is itself an integral domain, it is evident that the polynomial functions over an infinite integral domain form themselves an integral domain.

Since, then, $D \subseteq D[x]$ as well as $x \in D[x]$, it follows that $a_i x^i$ may be actually interpreted as a_i times x to the i -th power; the “superscript” of x in $D[x]$ is thus interchangeable with “power” of x .

Df. 4.1.2.5.11 A nonzero element $f(x)$ of the ring $R[x]$ is said to have *degree* n , denoted by $\deg f(x) = n$, if n is the largest nonnegative integer such that x^n has a nonzero coefficient in $f(x)$, which itself is called the *leading coefficient* of $f(x)$. The *zero polynomial*, $0x^k$, then, has no degree, hence no leading coefficient, while the nonzero coefficient of x^0 in $f(x)$ is called the *constant term* of $f(x)$.

Example:

$3x^2 + 2x - 1$, $5x$, 4 , of $P(x)$ over I , have degrees 2, 1, 0, leading coefficients 3, 5, 4, and constant terms $-1, 0, 4$ respectively.

Th. 4.1.2.5.12 If $f(x)$ and $g(x)$ are nonzero elements of the set $D[x]$ of all polynomials over an integral domain D , then

$$\deg(f(x)g(x)) = \deg f(x) + \deg g(x),$$

and

$$\deg(f(x) + g(x)) \leq \deg g(x)$$

if $\deg f(x) \leq \deg g(x)$ and $f(x) + g(x) \neq 0x^k$. (Cf. Prob. 8.)

$D[x]$ as such may be considered *quasi-partially ordered*, defined as follows.

Df. 4.1.2.5.13 For any $f(x), g(x) \in D[x]$, $f(x) < g(x)$ iff either $\deg f(x) < \deg g(x)$ or $f(x) = 0$ and $g(x) \neq 0$.

Example:

$$x^2 - 2 < -x^3, \quad 0 < -1, \text{ etc. in } D[x].$$

It is evident that $D[x]$ as such is not quite partially ordered, since ordering in $D[x]$ is neither reflexive nor symmetric; nor is it antisymmetric, although it is asymmetric and transitive. Neither does the trichotomy law hold for $D[x]$, since it cannot be said that $f(x) < g(x)$ or $f(x) = g(x)$ or $f(x) > g(x)$ when $f(x)$ and $g(x)$ are distinct and have the same degree.

Df. 4.1.2.5.14 A polynomial $f(x)$ is called *monic* if $\deg f(x) = n$ and the coefficient of x^n is 1.

Example:

$$x^7 - 2x + 1, \quad x + 3, \text{ etc. are monic polynomials.}$$

Th. 4.1.2.5.15 If $f(x), g(x) \in D[x]$ and $g(x)$ is monic, then there uniquely exist $q(x), r(x) \in D[x]$ such that

$$f(x) = g(x)q(x) + r(x), \quad 0 \leq r(x) < g(x)$$

(Cf. Prob. 9)

In analogy to q and r in Th. 4.1.2.3.13, $q(x)$ in this context may be called the *quotient* and $r(x)$ the *remainder* of $f(x)$ divided by $g(x)$.

Th. 4.1.2.5.16 (Remainder Theorem). For every $f(x) \in D[x]$ and $d \in D$,

$$f(x) = (x - d)q(x) + f(d)$$

where $f(d) = \sum_i a_i d^i = d' \in D$ if $f(x) = \sum_i a_i x^i$, $i = 0, 1, \dots, n$. (Cf. Prob. 12.)

The concept of “divisibility”, defined with respect to I (cf. Df. 4.1.2.3.9) may be carried over to $D[x]$, since Th. 4.1.2.5.16 may be rephrased as follows: $f(x)$ over D is divisible by $x - d$, $d \in D$, iff $f(d) = 0$.

Or, using the concept “factor” in the same context, the theorem may be further restated: $q(x) | f(x)$ (or $g(x) | f(x)$) iff $f(d) = 0$, $d \in D$.

The necessary and sufficient condition $f(d) = 0$ now yields the following definition.

Df. 4.1.2.5.17 If $f(d) = 0$ in Th. 4.1.2.5.16, then d is called a *root* (or *zero*) of $f(x)$.

This new term yields the *Factor Theorem*, the third restatement of Th. 4.1.2.5.16: For every $f(x) \in D[x]$, $x - d$ is a factor of $f(x)$ iff $d \in D$ is a root of $f(x)$.

This is generalized as follows:

Th. 4.1.2.5.18 If $f(x) \in D[x]$ and $\deg f(x) = n$, then $f(x)$ has at most n distinct roots in D . (Cf. Prob. 17.)

In general, many properties of the set of polynomials in a *single* indeterminate (or variable), as studied above, may be extended to the set of polynomials in *several* indeterminates.

Let $D[x]$ be, as before, an integral domain of polynomials in x over an integral domain D , and if an indeterminate y , independent from x , is to commute with all elements (i.e. polynomials) in $D(x)$, then $D[y[x]]$ (or $D[x][y]$) denotes the set of polynomials in y which have, as their coefficients, polynomials in x . As such, $D[y[x]]$ actually forms an integral domain (cf. Supplementary Problem 4.30) and is more simply written as $D[x, y]$, which initiates a further generalization $D[x_1, x_2, \dots, x_n]$, which ultimately denotes an integral domain of polynomials in n indeterminates x_1, x_2, \dots, x_n ; e.g. $f(x_1, x_2, \dots, x_n) \in D[x_1, x_2, \dots, x_n]$.

Th. 4.1.2.5.19 If $f(x_1, x_2, \dots, x_n) \in D[x_1, x_2, \dots, x_n]$, then it has the following form:

$$A_1 x_1^{a_1} x_2^{b_1} \dots x_n^{k_1} + \dots + A_m x_1^{a_m} x_2^{b_m} \dots x_n^{k_m}$$

where $A_i \in D$ and $a_i, b_i, \dots, k_i \in I^+$, $i = 1, 2, \dots, m$. (Cf. Prob. 18.)

As can be readily verified, $f(x_1, x_2, \dots, x_n) = 0$ iff each coefficient $A_i = 0$, and $f(x_1, x_2, \dots, x_n) = g(x_1, x_2, \dots, x_n)$ iff f and g have the same coefficients in D . Other properties of $D[x_1, x_2, \dots, x_n]$ can be obtained likewise (cf. Supplementary Problems 4.27-31).

Df. 4.1.2.5.20 The sum of the degrees of the indeterminates in a term is called its *dimension*, and a polynomial is called *homogeneous*, of dimension n , if every term of the polynomial has the same dimension n . (Cf. §5.2.1, Prob. 24, §5.2.2, Prob. 5, etc.)

Example:

$$3x - 2y + z, \quad x_1^2 + x_1 x_2 - x_2^2, \quad x^3 + xy^2 + z^3 - 3xyz$$

are homogeneous polynomials of dimension 1, 2, 3, respectively.

Solved Problems

1. Prove Th. 4.1.2.5.6.

PROOF:

R1 and **R6** are given by Df. 4.1.2.5.4-5 themselves, which in turn verify **R2** and **R5**.

Let $p(x), q(x), r(x) \in R[x]$, where $p(x) = \sum_i a_i x^i$, $q(x) = \sum_j b_j x^j$, $r(x) = \sum_k c_k x^k$, $i = 0, 1, 2, \dots, l$, $j = 0, 1, 2, \dots, m$, $k = 0, 1, 2, \dots, n$. Then:

R3: zero polynomial, i.e. $z(x) = \sum_i 0x^i$.

R4: $(p(x))^{-1} = -p(x)$

$$(\because p(x) + (-p(x)) = \sum_i a_i x^i + \sum_i (-a_i) x^i = \sum_i (a_i + (-a_i)) x^i = \sum_i 0x^i = z(x))$$

$$\begin{aligned} \mathbf{R7:} \quad p(x)(q(x)r(x)) &= \left(\sum_i a_i x^i\right) \left(\left(\sum_j b_j x^j\right) \left(\sum_k c_k x^k\right)\right) = \left(\sum_i a_i x^i\right) \left(\sum_r \left(\sum_{j+k=r} b_j c_k\right) x^r\right) \\ &= \sum_s \left(\sum_{i+r=s} \left(\sum_{j+k=r} a_i b_j c_k\right)\right) x^s = \sum_s \left(\sum_{r'+k=s} \left(\sum_{i+j=r'} a_i b_j c_k\right)\right) x^s \\ &= \left(\left(\sum_{i+j=r'} a_i b_j\right) x^{r'}\right) \left(\sum_k c_k x^k\right) = (p(x)q(x))r(x) \end{aligned}$$

$$\begin{aligned} \mathbf{R8:} \quad p(x)(q(x) + r(x)) &= \left(\sum_i a_i x^i\right) \left(\sum_j b_j x^j + \sum_k c_k x^k\right) = \left(\sum_{i+j=r} a_i (b_j + c_k)\right) x^r \\ &= \sum_r \left(\left(\sum_{i+j=r} a_i b_j\right) + \left(\sum_{i+k=r} a_i c_k\right)\right) x^r = \left(\sum_i a_i x^i\right) \left(\sum_j b_j x^j\right) + \left(\sum_i a_i x^i\right) \left(\sum_k c_k x^k\right) \\ &= p(x)q(x) + p(x)r(x) \end{aligned}$$

Likewise, $(p(x) + q(x))r(x) = p(x)r(x) + q(x)r(x)$.

Hence $R[x]$ is a ring.

2. Prove Th. 4.1.2.5.7.

PROOF:

Given $p(x)$ and $q(x)$ as in Prob. 1 above,

$$p(x)q(x) = \left(\sum_i a_i x^i\right) \left(\sum_j b_j x^j\right) = \sum_k \left(\sum_{i+j=k} a_i b_j\right) x^k = \sum_k \left(\sum_{j+i=k} b_j a_i\right) x^k = q(x)p(x)$$

Hence, with this multiplicative commutativity, $R[x]$ is a commutative ring.

3. If a ring R is with unity e , then the ring $R[x]$ of polynomials over R is also with unity.

PROOF:

The unity of $R[x]$ is ex^0 , since, for any $p(x) \in R[x]$,

$$(ex^0)p(x) = (ex^0) \left(\sum_k a_k x^k\right) = \sum_k (ea_k) x^{0+k} = \sum_k a_k x^k = p(x)$$

and likewise, $(p(x))ex^0 = p(x)$.

4. If a ring R is without zero-divisors, so is the ring $R[x]$ of polynomials over R .

PROOF:

Suppose that $R[x]$ has zero-divisors, which implies that there exist some $f(x), g(x) \in R[x]$ such that $f(x) \neq 0$, $g(x) \neq 0$, and yet $f(x) \cdot g(x) = 0$.

But, since $f(x) = \sum_i a_i x^i \neq 0$, $i = 0, 1, \dots, n$, and $g(x) = \sum_j b_j x^j \neq 0$, $j = 0, 1, \dots, m$, imply $a_n x^n \neq 0$ and $b_m x^m \neq 0$, which in turn imply $a_n b_m x^{n+m} \neq 0$ ($\because a_i, b_j \in D$). Hence, by Df. 4.1.2.5.5, $f(x)g(x) \neq 0$, contradicting the initial assumption: $f(x)g(x) = 0$. Hence it is not the case that $R[x]$ has zero-divisors.

Second proof (in terms of "degree"). Assume the same as above; then $\deg f(x) = n \neq 0$ and $\deg g(x) = m \neq 0$ (cf. Df. 4.1.2.5.11), which imply $\deg(f(x)g(x)) = n + m \neq 0$, contradicting the assumption that $f(x)g(x) = 0$ which implies $\deg f(x) = 0$ or $\deg g(x) = 0$ (cf. Prob. 8 note, below), i.e. $\deg(f(x)g(x)) \neq n + m$.

Hence it must be the case that $R[x]$ contains no zero-divisors.

5. Prove Th. 4.1.2.5.8.

PROOF:

By Prob. 1-2 above, the set $R[x]$ of polynomials over a commutative ring R is a commutative ring, and by Prob. 3 and 4, $R[x]$ over a ring R with unity and without zero-divisors is a ring with unity and without zero-divisors, which implies, by Prob. 1 of §4.1.2.2, that the cancellation law holds for both R and $R[x]$. Hence $R[x]$ over an integral domain R is also an integral domain.

Conversely, if $R[x]$ is an integral domain, then a complex $C[x]$, of $R[x]$, which consists of all polynomials of the form $c_1x^0, c_i \in R$, is a subdomain of $R[x]$, since $c_1x^0, c_2x^0 \in C(x)$ implies

$$c_1x^0 + c_2x^0 = (c_1 + c_2)x^0 \in C(x) \quad \text{and} \quad c_1x^0 \cdot c_2x^0 = (c_1c_2)x^{0+0} = (c_1c_2)x^0 \in C[x]$$

including $0x^0, 1x^0 \in C[x]$.

Furthermore, the mapping $F: F(c_i) \leftrightarrow c_ix^0$ is evidently 1-1. Hence the set R , to which any of c_i belongs, is an integral domain.

6. Prove Th. 4.1.2.5.10.

PROOF:

If $a(x), b(x) \in D[x]$ and $a(x) \neq b(x)$, by hypothesis, then $c(x) = a(x) - b(x) \neq 0$ and $\deg c(x) \geq 0$ (cf. Prob. 8 below). But Th. 4.1.2.5.18 (cf. Prob. 17 below) then implies that $c(x)$ has at most n roots, which implies that D has more than n elements, since D is infinite. Hence there must exist some $d \in D$ such that $c(d) \neq 0$, implying that $a(d) \neq b(d)$. Hence the mappings, of D into D , defined by $a(x)$ and $b(x)$ are distinct, completing the proof.

7. Th. 4.1.2.5.10 does not hold for a finite integral domain D' .**PROOF:**

Since D' is finite, let $d_1, d_2, \dots, d_n \in D'$ and construct $a(x), b(x), c(x) \in D'[x]$ such that

$$a(x) = (x - d_1)(x - d_2) \cdots (x - d_n) \quad \text{and} \quad b(x) = a(x) \cdot c(x)$$

where $c(x) \neq 1$. Then, for any $d_i \in D'$, $1 \leq i \leq n$,

$$a(d_i) = b(d_i) = 0$$

yielding the identical mappings, of D' into D' , by $a(x)$ and $b(x)$, which completes the proof.

Note. In general, $a(x) = b(x)$, i.e. $a(x) - b(x) = c(x) = \sum_i c_i x^i = 0$ if either all $c_i = 0$ or $a(x), b(x) \in D'[x]$ where D' is a finite integral domain. The latter follows from Th. 4.1.2.5.18, since, if D' is not finite, there will be some remaining values other than d_i , $1 \leq i \leq n$, of x for which $c(x) \neq 0$.

8. Prove Th. 4.1.2.5.12.

PROOF:

(i) Since, by hypothesis, $f(x) = \sum_i a_i x^i \neq 0$, $i = 0, 1, \dots, n$, and $g(x) = \sum_j b_j x^j \neq 0$, $j = 0, 1, \dots, m$,

it follows that $a_n \neq 0$ and $b_m \neq 0$, and that $a_n \cdot b_m \neq 0$ since $a_i, b_j \in D$. Hence $a_n b_m x^{n+m} \neq 0$, which is, by Df. 4.1.2.4.5 itself, the term of the highest degree in $f(x)g(x)$, and it immediately follows that $\deg(f(x)g(x)) = n + m = \deg f(x) + \deg g(x)$.

(ii) Since, by hypothesis, $\deg f(x) \leq \deg g(x)$ and $f(x) + g(x) \neq 0x^k$, it follows at once from Df. 4.1.2.4.4 that $n \leq m$ and $\deg(f(x) + g(x)) = m \leq \deg g(x) = m$.

Note. $\deg(f(x)g(x)) = \deg f(x) + \deg g(x)$ does not always hold unless, strictly, $f(x), g(x) \in D[x]$, since otherwise $a_n b_m$ can be zero without either factor being zero (cf. Th. 4.1.2.2.4). Thus in the residue class modulo 4, $f(x) = 3 + 2x$ and $g(x) = 5 + 2x^2$ give $\deg f(x) = 1$ and $\deg g(x) = 2$, yet $\deg(f(x)g(x)) = 2(\neq 3)$, since $f(x)g(x) = 15 + 10x + 6x^2$.

In $D[x]$, then, $\deg(f(x)g(x)) = 0$ iff $\deg f(x) = 0$ or $\deg g(x) = 0$, and $\deg(f(x) + g(x)) = 0$ if $\deg f(x) = 0$ and $\deg g(x) = 0$.

9. Prove Th. 4.1.2.5.15.

PROOF:

(i) If $g(x) = x^0$, then evidently

$$f(x) = g(x)f(x) + 0$$

i.e. $q(x) = f(x)$ and $r(x) = 0x^k$, proving the theorem.

(ii) If $f(x) < g(x)$, then again

$$f(x) = g(x) \cdot 0 + f(x), \quad 0 \leq f(x) < g(x)$$

where $q(x) = 0x^k$ and $r(x) = f(x)$, proving the theorem.

(iii) If $f(x) > g(x)$ or $\deg f(x) = \deg g(x)$, where $\deg g(x) \geq 1$, then proof is carried out by induction as follows.

Let $f(x) = \sum_i a_i x^i$, $i = 0, 1, \dots, m$, and $g(x) = \sum_j b_j x^j$, $j = 0, 1, \dots, n$, where the last term is x^n , by hypothesis; then, by the initial assumption, $m \geq n \geq 1$. Now assume that the theorem holds for every polynomial $f(x)$ of degree $k < m$, and let

$$f'(x) = f(x) - g(x) \cdot a_m x^{m-n} \quad (1)$$

Then $f'(x)$ has degree $k < m$, since the term of highest degree in $g(x) \cdot a_m x^{m-n}$ is $a_m x^m$, $g(x)$ being monic, which cancels $a_m x^m$ in $f(x)$. If, for some $q'(x), r(x) \in D[x]$,

$$f'(x) = g(x)q'(x) + r(x), \quad 0 \leq r(x) \leq g(x) \quad (2)$$

then, substituting (2) in (1),

$$f(x) = g(x) \cdot (q'(x) + a_m x^{m-n}) + r(x), \quad 0 \leq r(x) < g(x) \quad (3)$$

where $q(x) = q'(x) + a_m x^{m-n}$ and $r(x) = r(x)$, again proving the theorem.

Furthermore, the assumption that $q(x)$ and $r(x)$ are not unique, i.e.,

$$f(x) = g(x)q(x) + r(x), \quad 0 \leq r(x) < g(x),$$

and

$$f(x) = g(x)q'(x) + r'(x), \quad 0 \leq r'(x) < g(x)$$

will imply

$$(q'(x) - q(x))g(x) = r(x) - r'(x) \quad (4)$$

If $r(x) - r'(x) \neq 0$, then $(q'(x) - q(x))g(x) = h(x) \neq 0$, which, however, is impossible, since it implies $\deg h(x) \geq \deg g(x)$ while $\deg(r(x) - r'(x)) < \deg g(x)$. Hence it must be the case in (4) that $r(x) - r'(x) = 0$, in which case $r(x) = r'(x)$ and consequently $q(x) = q'(x)$, completing the proof.

10. Find the quotient and remainder in $D[x]$ of $2x^5 - 54x^3 + 21x^2 - 3x + 4$ divided by $x - 5$.

Solution:

By the "long division" process,

$$\begin{array}{r} 2x^4 + 10x^3 - 4x^2 + x + 2 \\ x - 5 \overline{) 2x^5 - 54x^3 + 21x^2 - 3x + 4} \\ \underline{2x^5 - 10x^4} \\ 10x^4 - 54x^3 \\ \underline{10x^4 - 50x^3} \\ -4x^3 + 21x^2 \\ \underline{-4x^3 + 20x^2} \\ x^2 - 3x \\ \underline{x^2 - 5x} \\ 2x + 4 \\ \underline{2x - 10} \\ 14 \end{array} \quad \begin{array}{l} = q(x) \\ = f(x) \\ \\ \\ \\ \\ \\ \\ \\ \\ = r(x). \end{array}$$

Or, more simply, by "synthetic division",

$$\begin{array}{r|rrrrrr} 2 & 0 & -54 & 21 & -3 & 4 & 5 \\ & 10 & 50 & -20 & 5 & 10 & \\ \hline 2 & 10 & -4 & 1 & 2 & 14 & \end{array}$$

In either case, $2x^5 - 54x^3 + 21x^2 - 3x + 4 = (x - 5)(2x^4 + 10x^3 - 4x^2 + x + 2) + 14$.

11. Express the following polynomial $a(x)$ in x as a polynomial in $x - 3$:

$$x^5 - 3x^4 + 2x^2 - 6x - 11$$

Solution:

Repeating synthetic divisions,

1	-3	0	2	-6	-11	<u>3</u>
	3	0	0	6	0	
1	0	0	2	0	-11	
	3	9	27	87		
1	3	9	29	87		
	3	18	81			
1	6	27	110			
1	3	27				
1	9	54				
	3					
1	12					

Hence $a(x) = (x-3)^5 + 12(x-3)^4 + 54(x-3)^3 + 110(x-3)^2 + 87(x-3) - 11$.

Second solution. Let $x-3 = y$; then $x = y+3$, and

$$\begin{aligned}
 a(x) &= a(y+3) = (y+3)^5 - 3(y+3)^4 + 2(y+3)^3 - 6(y+3) - 11 \\
 &= y^5 + 12y^4 + 54y^3 + 110y^2 + 87y - 11 \\
 &= (x-3)^5 + 12(x-3)^4 + 54(x-3)^3 + 110(x-3)^2 + 87(x-3) - 11
 \end{aligned}$$

Note. The reasoning behind the first method runs as follows: Let $x-3 = y$ and $a(x) = py^5 + qy^4 + ry^3 + sy^2 + ty + u$; then u is the remainder when $a(x)$ is divided by y , where the quotient, i.e. $a(x)/y$, is $py^4 + qy^3 + ry^2 + sy + t$, and t will be the remainder when $a(x)/y$ is further divided by y . Repeat the process, finding t, s, r, q, p successively.

12. Prove Th. 4.1.2.5.16.

PROOF:

Since, by Th. 4.1.2.4.13, there uniquely exist $q(x), r(x) \in D[x]$ such that

$$f(x) = (x-d)q(x) + r(x), \quad 0 \leq r(x) < (x-d)$$

it follows that either $r(x) = 0$ or $\deg r(x) = 0$, and that, in either case, $r(x) = d' \in D$, and $f(d) = (d-d)g(d) + d'$, i.e. $d' = f(d)$, completing the proof.

Second proof. Let $f(x) = \sum_i a_i x^i$ and $f(d) = \sum_i a_i d^i$, $i = 0, 1, \dots, n$; then, by the binomial formula,

$$f(x) - f(d) = \sum_i a_i x^i - \sum_i a_i d^i = \sum_i a_i (x^i - d^i) = \sum_i a_i ((x-d)(x^{n-1} + x^{n-2}d + \dots + d^{n-1}))$$

Hence $f(x) - f(d) = (x-d)q(x)$, where $q(x) = \sum_j b_j x^j$, $j = 0, 1, \dots, n-1$; or $f(x) = (x-d)q(x) + f(d)$.

13. If $(x-d) \mid f(x)$, where $f(x) = \sum_i a_i x^i$, $i = 0, 1, \dots, n$, then $d \mid a_0$.

PROOF:

By Th. 4.1.2.5.16, $f(d) = 0$, i.e.,

$$a_0 + a_1 d + \dots + a_n d^n = 0$$

i.e. $a_0 = -(a_1 d + \dots + a_n d^n) = -d(a_1 + \dots + a_n d^{n-1})$. Hence $d \mid a_0$.

14. Factor: $x^4 - 2x^2 + 3x - 2 = f(x)$.

Solution:

By Prob. 13, $d \mid 2$ if $(x-d) \mid f(x)$; hence $d = \pm 1$ or ± 2 . Checking, $f(1) = 1 - 2 + 3 - 2 = 0$ and $f(-2) = 16 - 8 - 6 - 2 = 0$. Hence $f(x) = (x-1)(x+2)g(x)$, and $g(x)$ is then found by the method of Prob. 11, i.e. by repeating the synthetic division:

1	0	-2	3	-2	<u>1</u>
	1	1	-1	2	
1	1	-1	2	0	<u>-2</u>
	-2	2	-2		
1	-1	1	0		

Hence $g(x) = x^2 - x + 1$, and $x^4 - 2x^2 + 3x - 2 = (x-1)(x+2)(x^2 - x + 1)$.

15. If $(x^2 - \epsilon^2) \mid (ax^4 + bx^3 + cx^2 + dx + e)$, where $\epsilon \neq 0$, then $ad^2 + b^2e = bcd$.

PROOF:

Let $f(x)$ denote the given polynomial; then, by hypothesis,

$$f(\epsilon) = a\epsilon^4 + b\epsilon^3 + c\epsilon^2 + d\epsilon + e \quad (1)$$

and

$$f(-\epsilon) = a\epsilon^4 - b\epsilon^3 + c\epsilon^2 - d\epsilon + e \quad (2)$$

Hence

$$a\epsilon^4 + c\epsilon^2 + e = 0 \quad (3)$$

and

$$b\epsilon^3 + d\epsilon = 0 \quad (4)$$

Since $\epsilon \neq 0$, it follows from (4) that $b\epsilon^2 = -d$ and $b^2\epsilon^4 = d^2$, which are substituted in (3) multiplied by b^2 , viz., $ad^2 - bcd + b^2e = 0$ or $ad^2 + b^2e = bcd$.

16. Find the necessary and sufficient condition that $(x-d)^2 \mid f(x)$, where

$$f(x) = \sum_i a_i x^i, \quad i = 0, 1, \dots, n$$

Solution:

By synthetic division,

$$\begin{array}{r|rrrr} a_n & a_{n-1} & a_{n-2} & \dots & \\ & a_n d & a_n d^2 + a_{n-1} d & \dots & \\ \hline a_n & a_n d + a_{n-1} & a_n d^2 + a_{n-1} d + a_{n-2} & \dots & \\ & a_1 & a_0 & & \\ & a_n d^{n-1} + a_{n-1} d^{n-2} + \dots + a_2 d & a_n d^n + a_{n-1} d^{n-1} + \dots + a_1 d & & \boxed{d} \\ \hline & a_n d^{n-1} + a_{n-1} d^{n-2} + \dots + a_2 d + a_1 & a_n d^n + a_{n-1} d^{n-1} + \dots + a_1 d + a_0 & & \end{array}$$

Hence $f(x) = (x-d)g(x) + f(d)$ where

$$g(x) = a_n x^{n-1} + (a_n d + a_{n-1})x^{n-2} + \dots + (a_n d^{n-1} + a_{n-1} d^{n-2} + \dots + a_2 d + a_1),$$

and the necessary and sufficient condition that $(x-d)^2 \mid f(x)$ is, then, $f(d) = 0$ and $g(d) = 0$, i.e.,

$$\begin{aligned} & a_n d^{n-1} + (a_n d + a_{n-1})d^{n-2} + \dots + (a_n d^{n-1} + a_{n-1} d^{n-2} + \dots + a_2 d + a_1) \\ &= n a_n d^{n-1} + (n-1)a_{n-1} d^{n-2} + \dots + a_1 = 0 \end{aligned}$$

17. Prove Th. 4.1.2.5.18.

PROOF:

(i) If $n = 0$, it is then evident that $f(x)$ has no root, satisfying the theorem.

(ii) If $n = 1$, say $f(x) = a_0 + a_1 x$, $a_1 \neq 0$, then $f(x)$ may or may not have a root. Thus $f(x) = 1 + 2x$, $x \in I$, has no root, while if it has a root d , then, by Th. 4.1.2.4.16,

$$f(x) = (x-d)g(x)$$

where, obviously, $g(x) = a_1$, and d is the only root, since if there exists another root, say $d' \neq d$, i.e. $d' - d \neq 0$, then

$$f(d') = (d' - d)g(x) \neq 0,$$

revealing that d' is not a root of $f(x)$. Hence $f(x)$ has at most one root if $\deg f(x) = 1$.

(iii) Assume that every polynomial of degree $k-1$, $k > 1$, has at most $k-1$ roots, and let $f(x)$ be any polynomial of degree k .

If $f(x)$ has no root, then the case is brought back to (i).

If $f(x)$ has a root d , then again, by Th. 4.1.2.5.16,

$$f(x) = (x-d)g(x),$$

where $\deg g(x) = k-1$ and, by the initial assumption, $g(x)$ has at most $n-1$ distinct roots. Furthermore, any root d' of $f(x)$ distinct from d is again a root of $g(x)$, since

$$f(d') = (d' - d)g(d') = 0$$

implies $g(d') = 0$. Hence $f(x)$ has at most k distinct roots.

In general, therefore, any polynomial of degree n has at most n distinct roots, completing the proof.

18. Prove Th. 4.1.2.5.19.

PROOF:

Consider, first, the set $D[x_1, x_2]$ of polynomials in two indeterminates x_1, x_2 , where each element is a polynomial in x_2 over $D[x_1]$; i.e. if $f(x_1, x_2) \in D[x_1, x_2]$, then

$$f(x_1, x_2) = a_0(x_1)x_2^0 + a_1(x_1)x_2^1 + \cdots + a_m(x_1)x_2^m \quad (1)$$

where each term is again a polynomial in x_1 over D , i.e.,

$$a_i(x_1) = a_{i_0} + a_{i_1}x_1^1 + \cdots + a_{i_{m_i}}x_1^{m_i}, \quad i = 0, 1, \dots, m, \text{ and } a_{ij} \in D, \quad j = 0, 1, \dots, i \quad (2)$$

Since $D[x_1] \subset D[x_1, x_2]$, where the operations in $D[x_1]$ are preserved in $D[x_1, x_2]$, (1) and (2) are now combined as follows:

$$f(x_1, x_2) = A_1x_1^{a_1}x_2^{b_1} + A_2x_1^{a_2}x_2^{b_2} + \cdots + A_mx_1^{a_m}x_2^{b_m} \quad (3)$$

where $A_i \in D$, $m, a_i, b_i \in I^+$, $i = 1, 2, \dots, m$, and there are no similar terms, i.e. terms different only by their coefficients, in (3). For, if $A_ix_1^{a_i}x_2^{b_i}$ and $A_jx_1^{a_j}x_2^{b_j}$ are similar, then $a_i = a_j$, $b_i = b_j$, and

$$A_ix_1^{a_i}x_2^{b_i} + A_jx_1^{a_j}x_2^{b_j} = (A_i + A_j)x_1^{a_i}x_2^{b_i}$$

combining two like terms into one.

Likewise, if $f(x_1, x_2, \dots, x_n) \in D[x_1, x_2, \dots, x_n]$, i.e. f is a polynomial in x_n over $D[x_1, x_2, \dots, x_{n-1}]$, then

$$f(x_1, x_2, \dots, x_n) = a_0(x_1, x_2, \dots, x_{n-1})x_n^0 + a_1(x_1, x_2, \dots, x_{n-1})x_n^1 + \cdots + a_m(x_1, x_2, \dots, x_{n-1})x_n^m \quad (4)$$

where $a_i(x_1, x_2, \dots, x_{n-1}) = A_{i_1}x_1^{p_1}x_2^{q_1} \cdots x_{n-1}^{t_1} + \cdots + A_{i_j}x_1^{p_j}x_2^{q_j} \cdots x_{n-1}^{t_j} \quad (5)$

and again, since $D[x_1, x_2, \dots, x_{n-1}] \subset D[x_1, x_2, \dots, x_n]$, where the operations in $D[x_1]$ are preserved, substitute (5) in (4), combining similar terms, and

$$f(x_1, x_2, \dots, x_n) = A_1x_1^{a_1}x_2^{b_1} \cdots x_n^{k_1} + \cdots + A_mx_1^{a_m}x_2^{b_m} \cdots x_n^{k_m} \quad (6)$$

where $A_i \in D$ and $a_i, b_i, \dots, k_i \in I^+$, $i = 1, 2, \dots, m$, completing the proof.

Note. It was presumed throughout the proof that $D[x_1, x_2, \dots, x_n]$ is an integral domain, which can be readily proved (cf. Supplementary Problems 4.27-4.31).

§4.1.3 Noncommutative Rings

4.1.3.1 SFIELDS and QUATERNIONS

Df. 4.1.3.1.1 A ring is called a *sfield* (or *skew field*, or *quasi-field*, or *noncommutative field*, or *division ring*), denoted by F^* , if it contains more than one element and, for every $a \in F^*$, $a \neq 0$, the equation $ax = b$ has a solution for any $b \in F^*$.

This definition suggests a reason for calling F^* a division ring, which also may be defined as a ring with unity and with a multiplicative inverse for nonzero elements (cf. Prob. 2-3).

For example, any set which forms a field forms *a fortiori* a sfield; as is obvious in Df. 4.1.3.1.1, any sfield is an integral domain, although not conversely, since the ring I of all integers is an integral domain, yet definitely not a division ring.

Sfields are, by definition, noncommutative; hence fields may be defined as follows:

Df. 4.1.3.1.2 A field is a commutative sfield.

In the language of Df. 4.1.2.4.1, a sfield is then a set which is a module, i.e. an additive Abelian group, with identity 0 and also a multiplicative group except for 0, satisfying distributive laws under addition. It is exemplified by the following definition.

Df. 4.1.3.1.3 The set \bar{Q} of *quaternions* (or *Hamilton quadruples*) consists of the elements of the form

$$\bar{a} = a_1 + a_2i + a_3j + a_4k$$

where the a_r , $r=1,2,3,4$, are real coefficients, which are commutative with i, j, k and which obey associative and distributive laws, and the following operative rules define i, j, k :

$$\begin{aligned} i^2 &= j^2 = k^2 = -1 \\ ij &= k, \quad jk = i, \quad ki = j \\ ji &= -k, \quad kj = -i, \quad ik = -j \end{aligned}$$

(cf. §3.2.4, Prob. 10, and §3.2.5, Prob. 23). The a_r are sometimes called the *coordinates* of a , which is then more simply denoted by a quadruple (a_1, a_2, a_3, a_4) , or even more simply: (a_r) , $r=1,2,3,4$.

Df. 4.1.3.1.4 Two quaternions $\bar{a} = (a_r)$ and $\bar{b} = (b_r)$, $r=1,2,3,4$, are equal iff $a_r = b_r$.

Df. 4.1.3.1.5 Addition and multiplication in \bar{Q} are defined as follows:

$$\begin{aligned} \bar{a} + \bar{b} &= (a_1 + a_2i + a_3j + a_4k) + (b_1 + b_2i + b_3j + b_4k) \\ &= (a_1 + b_1) + (a_2 + b_2)i + (a_3 + b_3)j + (a_4 + b_4)k \\ \bar{a} \cdot \bar{b} &= (a_1b_1 - a_2b_2 - a_3b_3 - a_4b_4) + (a_1b_2 + a_2b_1 + a_3b_4 - a_4b_3)i \\ &\quad + (a_1b_3 - a_2b_4 + a_3b_1 + a_4b_2)j + (a_1b_4 + a_2b_3 - a_3b_2 + a_4b_1)k \end{aligned}$$

and, for any real number c ,

$$c\bar{a} = \bar{a}c = ca_1 + ca_2i + ca_3j + ca_4k$$

The prescribed multiplication reveals at once the noncommutativity: $\bar{a} \cdot \bar{b} \neq \bar{b} \cdot \bar{a}$. (Cf. Prob. 5.)

Th. 4.1.3.1.6 Addition and multiplication in \bar{Q} are associative. (Cf. Prob. 4.)

Df. 4.1.3.1.7 The *conjugate* \bar{a}' of a quaternion $\bar{a} = a_1 + a_2i + a_3j + a_4k$ is defined to be $\bar{a}' = a_1 - a_2i - a_3j - a_4k$.

Manifestly, $\bar{a}' = 2a_1 - \bar{a}$, and the conjugate of the conjugate of \bar{a} is \bar{a} itself (cf. Prob. 6). Also, as can be readily verified (cf. Prob. 6), multiplicative commutativity holds for a quaternion and its conjugate, the product of which produces a real number, defined as follows:

Df. 4.1.3.1.8 The product of \bar{a} and \bar{a}' , called the *norm* of \bar{a} , is denoted by $N(\bar{a})$.

In general, the norm, a real number, of the product of two quaternions equals the product of their norms (cf. Prob. 9).

Th. 4.1.3.1.9 If $N(\bar{a}) \neq 0$, then \bar{a} has a multiplicative inverse \bar{a}^{-1} in \bar{Q} . (Cf. Prob. 10.)

This theorem prepares for establishing Q as a noncommutative field, viz.:

Th. 4.1.3.1.10 The set Q of quaternions forms a sfield. (Cf. Prob. 11.)

Because of the noncommutativity, the automorphism of Q takes a specific form: $\bar{a} \leftrightarrow \bar{a}'$ and $\bar{b} \leftrightarrow \bar{b}'$, where $\bar{a}, \bar{a}', \bar{b}, \bar{b}' \in \bar{Q}$, imply $\bar{a}\bar{b} \leftrightarrow \bar{b}'\bar{a}'$ (cf. Prob. 12). Such an automorphism is called an *anti- (or reciprocal) automorphism*.

In analogy to the absolute values of complex numbers (cf. Df. 5.1.3.5), their counterpart in \bar{Q} may be also defined as follows:

Df. 4.1.3.1.11 The *absolute value* of a quaternion \bar{a} is denoted by $|\bar{a}|$, representing

$$|\bar{a}| = |a_1 + a_2i + a_3j + a_4k| = \sqrt{a_1^2 + a_2^2 + a_3^2 + a_4^2}$$

(Cf. Supplementary Prob. 4.33.)

Solved Problems

1. A sfield F^* has no proper divisor of zero.

PROOF:

Suppose that $a, b \in F^*$, where $a \neq 0$, $b \neq 0$; then, by Df. 4.1.2.3.1, there exists $x \in F^*$ such that $ax = b$ and also $y \in F^*$ such that $x = by$, i.e.,

$$aby = ax = b$$

which implies, by the same definition, $ab \neq 0$, proving that a product of nonzero elements of F^* cannot be zero, or what is the same, that the vanishing of a product in F^* entails at least one zero factor.

2. A sfield F^* is a ring with unity.

PROOF:

Let $a, e \in F^*$, $a \neq 0$, $e \neq 0$, such that $ae = a$; then $ae^2 = (ae)e = ae$, which implies $e^2 = e$.

Furthermore, for any element b of F^* , $(b - be)e = 0$ and $(b - eb)e = 0$, which implies $be = eb = b$, proving that e is the unity of F^* .

3. Every nonzero element of F^* has a multiplicative inverse.

PROOF:

By Df. 4.1.2.3.1, there exists $x \in F^*$ such that $ax = e$, where $x \neq 0$ and $a, e \in F^*$. Then, by the same definition,

$$(xa - e)x = xax - ex = x(ax) - ex = xe - ex = 0$$

which implies $xa - e = 0$, i.e. $xa = e$, proving that x is the multiplicative inverse a^{-1} of a .

4. Addition in \bar{Q} is both associative and commutative.

PROOF:

If $\bar{a}, \bar{b}, \bar{c} \in \bar{Q}$, then

$$\begin{aligned} \bar{a} + (\bar{b} + \bar{c}) &= (a_1 + a_2i + a_3j + a_4k) + ((b_1 + c_1) + (b_2 + c_2)i + (b_3 + c_3)j + (b_4 + c_4)k) \\ &= (a_1 + b_1 + c_1) + (a_2 + b_2 + c_2)i + (a_3 + b_3 + c_3)j + (a_4 + b_4 + c_4)k \\ &= ((a_1 + b_1) + (a_2 + b_2)i + (a_3 + b_3)j + (a_4 + b_4)k) + (c_1 + c_2i + c_3j + c_4k) = (\bar{a} + \bar{b}) + \bar{c} \\ \text{and } \bar{a} + \bar{b} &= (a_1 + b_1) + (a_2 + b_2)i + (a_3 + b_3)j + (a_4 + b_4)k \\ &= (b_1 + a_1) + (b_2 + a_2)i + (b_3 + a_3)j + (b_4 + a_4)k = \bar{b} + \bar{a} \end{aligned}$$

5. Multiplication in \bar{Q} is associative, but not commutative.

PROOF:

If $\bar{a}, \bar{b}, \bar{c} \in \bar{Q}$, then

$$\begin{aligned} \bar{a}(\bar{b}\bar{c}) &= (a_1 + a_2i + a_3j + a_4k)((b_1c_1 - b_2c_2 - b_3c_3 - b_4c_4) + (b_1c_2 + b_2c_1 + b_3c_4 - b_4c_3)i \\ &\quad + (b_1c_3 - b_2c_4 + b_3c_1 + b_4c_2)j + (b_1c_4 + b_2c_3 - b_3c_2 + b_4c_1)k) \\ &= (a_1(b_1c_1 - b_2c_2 - b_3c_3 - b_4c_4) - a_2(b_1c_2 + b_2c_1 + b_3c_4 - b_4c_3) \\ &\quad - a_3(b_1c_3 - b_2c_4 + b_3c_1 + b_4c_2) - a_4(b_1c_4 + b_2c_3 - b_3c_2 + b_4c_1)) \\ &\quad + (a_1(b_1c_2 + b_2c_1 + b_3c_4 - b_4c_3) + a_2(b_1c_1 - b_2c_2 - b_3c_3 - b_4c_4) \\ &\quad + a_3(b_1c_4 + b_2c_3 - b_3c_2 + b_4c_1) - a_4(b_1c_3 - b_2c_4 + b_3c_1 + b_4c_2))i \\ &\quad + (a_1(b_1c_3 - b_2c_4 + b_3c_1 + b_4c_2) - a_2(b_1c_4 + b_2c_3 - b_3c_2 + b_4c_1) \\ &\quad + a_3(b_1c_1 - b_2c_2 - b_3c_3 - b_4c_4) + a_4(b_1c_2 + b_2c_1 + b_3c_4 - b_4c_3))j \\ &\quad + (a_1(b_1c_4 + b_2c_3 - b_3c_2 + b_4c_1) + a_2(b_1c_3 - b_2c_4 + b_3c_1 + b_4c_2) \\ &\quad - a_3(b_1c_2 + b_2c_1 + b_3c_4 - b_4c_3) + a_4(b_1c_1 - b_2c_2 - b_3c_3 - b_4c_4))k \\ &= ((a_1b_1 - a_2b_2 - a_3b_3 - a_4b_4) + (a_1b_2 + a_2b_1 + a_3b_4 - a_4b_3)i + (a_1b_3 - a_2b_4 + a_3b_1 + a_4b_2)j \\ &\quad + (a_1b_4 + a_2b_3 - a_3b_2 + a_4b_1)k)(c_1 + c_2i + c_3j + c_4k) \\ &= (\bar{a}\bar{b})\bar{c} \\ \text{but } \bar{b}\bar{a} &= (b_1a_1 - b_2a_2 - b_3a_3 - b_4a_4) + (b_1a_2 + b_2a_1 + b_3a_4 - b_4a_3)i \\ &\quad + (b_1a_3 - b_2a_4 + b_3a_1 + b_4a_2)j + (b_1a_4 + b_2a_3 - b_3a_2 + b_4a_1)k \\ &= (a_1b_1 - a_2b_2 - a_3b_3 - a_4b_4) + (a_1b_2 + a_2b_1 - a_3b_4 + a_4b_3)i \\ &\quad + (a_1b_3 + a_2b_4 - a_3b_1 - a_4b_2)j + (a_1b_4 + a_2b_3 - a_3b_2 + a_4b_1)k \\ &\neq \bar{a}\bar{b} \end{aligned}$$

Note. The noncommutativity is actually an immediate result from $ij \neq ji$, $jk \neq kj$, $ki \neq ik$.

6. Prove (i) $\bar{a}' = 2a_1 - \bar{a}$, (ii) $(\bar{a}')' = \bar{a}$, (iii) $\bar{a}\bar{a}' = \bar{a}'\bar{a}$.

PROOF:

$$\begin{aligned}
 \text{(i)} \quad & \text{Since } \bar{a} = a_1 + a_2i + a_3j + a_4k \text{ and } \bar{a}' = a_1 - a_2i - a_3j - a_4k, \\
 & 2a_1 - \bar{a}' = 2a_1 - (a_1 - a_2i - a_3j - a_4k) = a_1 + a_2i + a_3j + a_4k = \bar{a} \\
 \text{(ii)} \quad & (\bar{a}')' = (a_1 - a_2i - a_3j - a_4k)' = a_1 - (-a_2)i - (-a_3)j - (-a_4)k = \bar{a} \\
 \text{(iii)} \quad & \bar{a}\bar{a}' = (a_1 + a_2i + a_3j + a_4k)(a_1 - a_2i - a_3j - a_4k) \\
 & = (a_1a_1 + a_2a_2 + a_3a_3 + a_4a_4) + (a_1(-a_2) + a_2a_1 + a_3(-a_4) - a_4(-a_3))i \\
 & \quad + (a_1(-a_3) - a_2(-a_4) + a_3a_1 + a_4(-a_2))j + (a_1(-a_4) - a_2(-a_3) + a_3(-a_2) + a_4a_1)k \\
 & = a_1^2 + a_2^2 + a_3^2 + a_4^2
 \end{aligned}$$

$$\text{Likewise } \bar{a}'\bar{a} = a_1^2 + a_2^2 + a_3^2 + a_4^2.$$

$$\text{Hence } \bar{a}\bar{a}' = \bar{a}'\bar{a}.$$

7. $(\bar{a}\bar{b})' = \bar{b}'\bar{a}'$.

PROOF:

$$\begin{aligned}
 \bar{b}'\bar{a}' &= (b_1 - b_2i - b_3j - b_4k)(a_1 - a_2i - a_3j - a_4k) \\
 &= (b_1a_1 - (-b_2)(-a_2) - (-b_3)(-a_3) - (-b_4)(-a_4)) \\
 & \quad + (b_1(-a_2) + (-b_2)a_1 + (-b_3)(-a_4) - (-b_4)(-a_3))i \\
 & \quad + (b_1(-a_3) - (-b_2)(-a_4) + (-b_3)a_1 + (-b_4)(-a_2))j \\
 & \quad + (b_1(-a_4) + (-b_2)(-a_3) - (-b_3)(-a_2) + (-b_4)a_1)k \\
 &= (b_1a_1 - b_2a_2 - b_3a_3 - b_4a_4) - (b_1a_3 + b_2a_1 - b_3a_4 + b_4a_3)i \\
 & \quad - (b_1a_4 + b_2a_3 + b_3a_1 - b_4a_2)j - (b_1a_2 - b_2a_4 + b_3a_2 + b_4a_1)k \\
 &= (a_1b_1 - a_2b_2 - a_3b_3 - a_4b_4) + (-a_1b_3 + a_2b_1 + a_3b_4 - a_4b_2)i \\
 & \quad + (-a_1b_4 - a_2b_3 + a_3b_1 + a_4b_2)j + (-a_1b_2 + a_2b_4 - a_3b_2 + a_4b_1)k \\
 &= (\bar{a}\bar{b})'
 \end{aligned}$$

8. $N(\bar{a}\bar{b}) = N(\bar{a})N(\bar{b})$.

PROOF:

Applying Df. 4.1.3.1.8 twice, Prob. 5, Prob. 7, and Df. 4.1.3.1.5 successively,

$$N(\bar{a}\bar{b}) = (\bar{a}\bar{b})(\bar{a}\bar{b})' = \bar{a}\bar{b}\bar{b}'\bar{a}' = \bar{a}N(\bar{b})\bar{a}' = \bar{a}\bar{a}'N(b) = N(a)N(b)$$

9. $\bar{a}\bar{b} = 0$ implies $\bar{a} = 0$ or $\bar{b} = 0$.

PROOF:

Multiplying both sides of the given equation $\bar{a}\bar{b}$,

$$\bar{a}\bar{b}(\bar{a}\bar{b})' = 0(\bar{a}\bar{b})' = 0$$

while, by Prob. 8, $\bar{a}\bar{b}(\bar{a}\bar{b})' = N(\bar{a}\bar{b}) = N(\bar{a})N(\bar{b})$. Hence

$$N(\bar{a})N(\bar{b}) = 0$$

which implies $N(\bar{a}) = 0$ or $N(\bar{b}) = 0$, since both $N(\bar{a})$ and $N(\bar{b})$ are real numbers. But if $N(\bar{a}) = 0$, then $a_1^2 + a_2^2 + a_3^2 + a_4^2 = 0$, which implies $a_1 = a_2 = a_3 = a_4 = 0$, yielding $\bar{a} = 0$.

Likewise $\bar{b} = 0$ if $N(\bar{a}) \neq 0$ and $N(\bar{b}) = 0$, completing the proof.

10. Prove Th. 4.1.3.1.9.

PROOF:

Since the multiplicative identity in \bar{Q} is 1, i.e. $1 + i + j + k$ or $(1,1,1,1)$, and, by Df. 4.1.3.1.8 and Prob. 6, (iii), $N(\bar{a}) = \bar{a}\bar{a}' = \bar{a}'\bar{a}$, it follows that $\bar{a} \neq 0$ implies

$$\bar{a}(\bar{a}'/N(\bar{a})) = \bar{a}'(\bar{a}/N(\bar{a})) = N(\bar{a})/N(\bar{a}) = 1$$

Hence $\bar{a}^{-1} = \bar{a}'/N(\bar{a})$, completing the proof.

11. Prove Th. 4.1.3.1.10.

PROOF:

 \bar{Q} is a ring, since it satisfies **R1-8** as follows:

$$\bar{Q}1 \equiv \mathbf{R}1, \text{ by Df. 4.1.3.1.5.}$$

$$\bar{Q}2 \equiv \mathbf{R}2, \text{ by Prob. 4.}$$

$$\bar{Q}3 \equiv \mathbf{R}3, \text{ since } 0 + 0i + 0j + 0k = (0, 0, 0, 0) = 0.$$

$$\bar{Q}4 \equiv \mathbf{R}4, \text{ since } \bar{a} + (-\bar{a}) = (a_r) + (-a_r) = (a_r - a_r) = 0 = (-\bar{a}) + \bar{a}.$$

$$\bar{Q}5 \equiv \mathbf{R}5, \text{ by Prob. 4.}$$

$$\bar{Q}6 \equiv \mathbf{R}6, \text{ by Df. 4.1.3.1.5.}$$

$$\bar{Q}7 \equiv \mathbf{R}7, \text{ by Prob. 5.}$$

$$\begin{aligned}
\bar{Q}8 \equiv \mathbf{R}8, \text{ since } \bar{a}(\bar{b} + \bar{c}) &= (a_1 + a_2i + a_3j + a_4k)((b_1 + c_1) + (b_2 + c_2)i + (b_3 + c_3)j + (b_4 + c_4)k) \\
&= (a_1(b_1 + c_1) - a_2(b_2 + c_2) - a_3(b_3 + c_3) - a_4(b_4 + c_4)) \\
&\quad + (a_1(b_2 + c_2) + a_2(b_1 + c_1) + a_3(b_4 + c_4) - a_4(b_3 + c_3))i \\
&\quad + (a_1(b_3 + c_3) - (a_2(b_4 + c_4) + a_3(b_1 + c_1) + a_4(b_2 + c_2))j \\
&\quad + (a_1(b_4 + c_4) + a_2(b_3 + c_3) - (a_3(b_2 + c_2) + a_4(b_1 + c_1))k \\
&= ((a_1b_1 - a_2b_2 - a_3b_3 - a_4b_4) + (a_1b_2 + a_2b_1 + a_3b_4 - a_4b_3)i \\
&\quad + (a_1b_3 - a_2b_4 + a_3b_1 + a_4b_2)j + (a_1b_4 + a_2b_3 - a_3b_2 + a_4b_1)k \\
&\quad + ((a_1c_1 - a_2c_2 - a_3c_3 - a_4c_4) + (a_1c_2 + a_3c_1 + a_3c_4 - a_4c_3)i \\
&\quad + (a_1c_3 - a_2c_4 + a_3c_1 + a_4c_2)j + (a_1c_4 + a_2c_3 - a_3c_2 + a_4c_1)k) \\
&= \bar{a}\bar{b} + \bar{a}\bar{c}
\end{aligned}$$

$$\text{and likewise } (\bar{a} + \bar{b})\bar{c} = \bar{a}\bar{c} + \bar{b}\bar{c}.$$

Furthermore, by Prob. 10, \bar{Q} satisfies **F9-10** although it fails to yield **F11** (cf. Prob. 5 above). Hence \bar{Q} is a noncommutative field, i.e. sfield, completing the proof.

12. Set up a 1-1 correspondence between the set \bar{Q} of quaternions and the set \bar{Q}' of their conjugates.

Solution:

Since $\bar{a}, \bar{b} \in \bar{Q}$ and $\bar{a}', \bar{b}' \in \bar{Q}$ imply a correspondence

$$\bar{a} \leftrightarrow \bar{a}' \quad \text{and} \quad \bar{b} \leftrightarrow \bar{b}',$$

it follows from Df. 4.1.3.1.5, 7 and Prob. 5, 7 that

$$\bar{a} + \bar{b} \leftrightarrow \bar{a}' + \bar{b}' \quad \text{and} \quad \bar{a} \cdot \bar{b} \leftrightarrow \bar{b}' \cdot \bar{a}' \quad (\neq \bar{a}' \cdot \bar{b}')$$

Since obviously $\bar{Q}' \subset \bar{Q}$, the isomorphism between \bar{Q} and \bar{Q}' is an automorphism (which, more specifically, embodies an *antiautomorphism* through the multiplicative noncommutativity exemplified as above).

4.1.3.2 MATRICES

Df. 4.1.3.2.1 *Vectors* (or *hypercomplex numbers*) over a ring R , denoted by α, β, \dots , are the n -tuples (or n -uples) of elements a_i, b_i, \dots , $i = 1, 2, \dots, n$, of R , viz.,

$$\alpha = (a_1, a_2, \dots, a_n) = (a_i), \quad \beta = (b_1, b_2, \dots, b_n) = (b_i), \quad \dots$$

where $\alpha = \beta$ iff $a_i = b_i$, and for every $r \in R$,

$$\begin{aligned}
r\alpha &= (ra_1, ra_2, \dots, ra_n) = (ra_i), \quad \text{called scalar multiplication, and} \\
\alpha + \beta &= (a_1 + b_1, a_2 + b_2, \dots, a_n + b_n) = (a_i + b_i), \quad \text{called vector addition.}
\end{aligned}$$

Each of a_i is called a *component* of α , and since n is explicitly finite in this context, the n -tuples are called *finite vectors of order n* . A set of such vectors is called a *vector* (or *linear*) *space over R* , denoted by $V(R)$, or more explicitly $V_n(R)$, which is then called an *n -dimensional vector* (or *linear*) *space* (cf. Df. 4.1.3.2.7 below) if it satisfies the following theorem.

Th. 4.1.3.2.2 A vector (or linear) space over a field F , denoted by $V(F)$ or more simply V , is a module, i.e. an additive Abelian group, viz.,
V1-5 \equiv **R1-5**,

with the following additional properties:

V6. For every $a \in F$ and $\alpha \in V$, $a\alpha \in V$.

V7. For every $a, b \in F$ and $\alpha, \beta \in V$, (i) $a(\alpha + \beta) = a\alpha + a\beta$, (ii) $(a + b)\alpha = a\alpha + b\alpha$.

V8. For every $a, b \in F$ and $\alpha \in V$, (i) $(ab)\alpha = a(b\alpha)$, (ii) $1\alpha = \alpha$. (Cf. Prob. 1.)

Example:

$F(F)$, i.e. a field F over itself, obviously satisfies **V1-8** and is certainly a vector space; so is $\bar{R}(R)$, i.e. the set \bar{R} of real numbers over the set R of rational numbers, as can be readily verified (cf. Prob. 3).

Also, \bar{Q} , the set of quaternions, is obviously a vector space over the field \bar{R} of real numbers, in which sense it may be represented by $\bar{Q}(\bar{R})$; its elements, quaternions, are then regarded as vectors (or hypercomplex numbers) of order 4, or 4-dimensional vectors. Real and complex numbers (cf. Df. 5.1.2.10 and Df. 5.1.3.1), then, may be considered 1-dimensional and 2-dimensional hypercomplex numbers (i.e. vectors) respectively, evidently satisfying Df. 4.1.3.2.1.

As in other algebras (of groups, rings, etc.), a *subspace* of a vector space V is a complex of V , which is itself a vector space, satisfying Df. 4.1.3.2.1; it is articulated by the following theorem.

Th. 4.1.3.2.3 A complex U of a vector space V is a subspace of V iff $a\alpha + b\beta \in U$ for every $a, b \in F$ and $\alpha, \beta \in U$. (Cf. Prob. 5.)

Example:

$V_n(F)$ in itself is a subspace of $V_n(F)$; so is the zero subspace, the sole element of which is the *zero vector*, $o = (O_1, O_2, \dots, O_n)$. As is obvious in this example, the meet $U \cap W$ of any two subspaces U and W of a vector space V is again a subspace of V (cf. Prob. 8); so is their *linear sum*, defined by $\{\alpha + \beta\}$ for any $\alpha \in U$, $\beta \in W$ (cf. Prob. 8).

Df. 4.1.3.2.4 A *linear combination* in $V_n(F)$ is a sum

$$\alpha = c_1\alpha_1 + c_2\alpha_2 + \dots + c_k\alpha_k = \sum_i c_i\alpha_i, \quad i = 1, 2, \dots, k, \text{ where } 1 \leq k \leq n$$

for every $c_i \in F$ and $\alpha_i \in V_n(F)$. The vectors α_i are called *linearly independent* over F iff $\alpha = 0$ implies $c_i = 0$ for every $c_i \in F$; otherwise, they are called *linearly dependent* (cf. Df. 5.3.1.11).

Example:

$\alpha_1 = (1, 0, 0)$, $\alpha_2 = (0, 1, 0)$, $\alpha_3 = (0, 0, 1)$ in $V_3(\bar{R})$ are linearly independent, since $c_1(1, 0, 0) + c_2(0, 1, 0) + c_3(0, 0, 1) \neq (0, 0, 0)$ for any nonzero $c_1, c_2, c_3 \in \bar{R}$, while $\beta_1 = (2, 1, 1)$, $\beta_2 = (1, -1, 1)$, $\beta_3 = (5, 4, 2)$, say, are linearly dependent, since $3(2, 1, 1) - (1, -1, 1) - (5, 4, 2) = (0, 0, 0)$.

Th. 4.1.3.2.5 The set S of all linear combinations of any set of vectors α_i in a vector space $V_n(F)$ is a subspace of $V_n(F)$ and is called the *subspace spanned* (or *generated*) by α_i . (Cf. Prob. 7.)

Example:

The subspace spanned by a single nonzero vector α_1 is the set S_1 of all scalar products $c\alpha_1$, which may be geometrically represented by the straight line through the origin; the subspace spanned by two non-collinear vectors α_1 and α_2 is the plane through α_1, α_2 , and the origin.

Th. 4.1.3.2.6 The vectors $\alpha_1, \alpha_2, \dots, \alpha_k \in V_n(F)$, $1 \leq k \leq n$, are linearly dependent iff some α_i , $i = 1, 2, \dots, k$, is in the subspace S spanned by the remaining vectors. (Cf. Prob. 9.)

Stated otherwise: any set of n vectors in $V_m(F)$, $n > m$, is linearly dependent; hence the following definition.

Df. 4.1.3.2.7 The set of vectors ϵ_i , $1 \leq i \leq n$, is a *basis* of the subspace S of $V_n(F)$ if it forms a linearly independent set which spans S , and the *dimension* of S is the number of elements in a basis of S .

Example:

$\epsilon_1 = (1, 0, 0, \dots, 0)$, $\epsilon_2 = (0, 1, 0, \dots, 0)$, \dots , $\epsilon_n = (0, 0, 0, \dots, 1)$ form a basis of $V_n(F)$, the dimension of which is n , while $(1, 0)$ and $(0, 1)$, say, form a basis of a subspace $V_2(F)$ of $V_n(F)$; the zero subspace is considered to have no basis, since the zero vector alone is a linearly dependent set. In general, any two bases of a finite dimensional vector space have the same number of elements (cf. Prob. 10).

The concept of the n -tuples ordered in one way in V is now expanded to those ordered in two ways, viz.:

Df. 4.1.3.2.8 A rectangular array of *elements* (or *entries*, or *coordinates*) of a field F , having m rows and n columns, is called an m by n *matrix* A over F , denoted either compactly by $A = (a_{ij})$, $i = 1, 2, \dots, m$, $j = 1, 2, \dots, n$, or more explicitly,

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \quad (1)$$

where $a_{ij} \in F$, which designates the element in the i -th row and the j -th column and is sometimes called (i, j) -th element of M .

A as such may represent m vectors $\alpha_1, \alpha_2, \dots, \alpha_m$ of $V_n(F)$, i.e. the elements $a_{i1}, a_{i2}, \dots, a_{in}$ of the i -th row of A corresponding to the components $a_{i1}, a_{i2}, \dots, a_{in}$ of α_i . Or more directly, a single row

$$[a_{i1} \ a_{i2} \ \dots \ a_{in}] \quad (2)$$

is itself a 1 by n matrix, and a single column

$$\begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix} \quad (3)$$

is an m by 1 matrix, called a *row matrix* (or *row vector*) and a *column matrix* (or *column vector*) respectively and sometimes referred to simply as *vectors*.

Note. The brackets in (1), (2), (3) may be replaced by parentheses or double straight lines on each side of the array.

Df. 4.1.3.2.9 Two m by n matrices $A = (a_{ij})$ and $B = (b_{ij})$, $i = 1, 2, \dots, m$, $j = 1, 2, \dots, n$, are equal iff $a_{ij} = b_{ij}$.

If $m = 1$, the case is reduced to two row matrices, or vectors of order n , where $(a_{1j}) = (b_{1j})$, $j = 1, 2, \dots, n$, iff $a_{1j} = b_{1j}$ (cf. Df. 4.1.3.2.1); likewise, $(a_{i1}) = (b_{i1})$ implies $a_{i1} = b_{i1}$, $i = 1, 2, \dots, m$, and conversely, if $n = 1$, i.e. A and B are column matrices, or vectors of order m .

Df. 4.1.3.2.10 The vector addition and the scalar multiplication of Df. 4.1.3.2.1 hold for matrices; viz., given two m by n matrices $\alpha = (a_{ij})$, $\beta = (b_{ij})$, and any element $c \in F$, it is defined that

$$\begin{aligned} c\alpha &= c(a_{ij}) = (ca_{ij}), \text{ called scalar multiplication, and} \\ \alpha + \beta &= (a_{ij}) + (b_{ij}) = (a_{ij} + b_{ij}), \text{ called vector addition.} \end{aligned}$$

Example:

If $c = 2$, and if A and B are 2 by 3 matrices, say,

$$A = \begin{bmatrix} 3 & 2 & -1 \\ 4 & -3 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & -2 & 7 \\ 3 & 2 & -1 \end{bmatrix},$$

$$\text{then } cA = 2A = \begin{bmatrix} 6 & 4 & -2 \\ 8 & -6 & 2 \end{bmatrix} \quad \text{and} \quad A + B = \begin{bmatrix} 3+1 & 2-2 & -1+7 \\ 4+3 & -3+2 & 1-1 \end{bmatrix} = \begin{bmatrix} 4 & 0 & 6 \\ 7 & -1 & 0 \end{bmatrix}$$

Hence the following theorem, which is an immediate result from Df. 4.1.3.2.10, articulating the relation between matrices and vectors:

Th. 4.1.3.2.11 The set \bar{M} of all m by n matrices over F is a vector space over F under addition and scalar multiplication. (Cf. Prob. 12.)

This theorem is in fact intuitively evident from the considerations of row and column matrices in Df. 4.1.3.2.8.

Also, as can be easily verified (cf. Prob. 11), addition is both associative and commutative in \bar{M} , while multiplication, defined as follows, is not.

Df. 4.1.3.2.12 If a matrix B has the same number of rows as a matrix A has columns, then B is said to be *conformable* with respect to A (and in general not conversely), yielding their (*matrix*) *product* AB (and in general not BA), defined by the following row by column multiplication, called *matrix multiplication*:

$$AB = (a_{ij})(b_{ij}) = \left(\sum_k a_{ik} b_{kj} \right), \quad i=1,2,\dots,p \quad j=1,2,\dots,r$$

where the summation $a_{ik} b_{kj}$, k going from 1 to q , is feasible as A is a p by q matrix and B is a q by r matrix, i.e. B is conformable with respect to A .

It follows at once that A is not conformable with respect to B in this context, that is, unless it happens that $p=r$; even then, it seldom if ever follows that $AB=BA$, since it does not always follow that $\sum_k a_{ik} b_{kj} = \sum_k b_{ik} a_{kj}$ even if $p=q=r$ (cf.

Df. 4.1.3.2.13). Commutativity, then, does not belong to matrix multiplication, although associativity does if conformability is assured (cf. Prob. 17).

Example:

If $A = (a_{ij})$ is a 3 by 2 matrix, $B = (b_{ij})$ a 2 by 2 matrix, then

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}, \quad B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}, \quad \text{and} \quad AB = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \\ a_{31}b_{11} + a_{32}b_{21} & a_{31}b_{12} + a_{32}b_{22} \end{bmatrix}$$

while, because of conformability, BA cannot even exist in this case. Since AB is here a 3 by 2 matrix, however, multiplication can be further carried out likewise, yielding $(AB)C$, if C is a 2 by n matrix, for any natural number n ; the product ABC is then a 3 by n matrix (cf. Prob. 17).

If A is a 1 by n matrix and B an n by 1 matrix, their product AB is a 1 by 1 matrix, viz.,

$$AB = [a_{11} \ a_{12} \ \dots \ a_{1n}] \begin{bmatrix} b_{11} \\ b_{21} \\ \vdots \\ b_{n1} \end{bmatrix} = [a_{11}b_{11} + a_{12}b_{21} + \dots + a_{1n}b_{n1}]$$

which is called the *inner* (or *scalar* or *dot*) *product* of A and B as vectors; as such, it may be written more simply as $AB = (a_i b_i)$ for $A = (a_i)$ and $B = (b_i)$, $i=1,2,\dots,n$.

Df. 4.1.3.2.13 A matrix is said to be *square* if it is n by n , having the same number of rows and columns.

Df. 4.1.3.2.14 The *transpose* of an m by n matrix $A = (a_{ij})$, denoted by A^T , is the n by m matrix $B = (b_{ij})$, where $b_{ij} = a_{ji}$, $i = 1, 2, \dots, m$, $j = 1, 2, \dots, n$.

Example:

If A is a 2 by 3 matrix, then

$$A = (a_{ij}) = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \quad \text{and} \quad A^T = (a_{ji}) = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{13} & a_{23} \end{bmatrix}$$

Th. 4.1.3.2.15 If two matrices A and B are conformable with respect to each other, then $(AB)^T = B^T A^T$. (Cf. Prob. 22.)

Df. 4.1.3.2.16 The elements a_{ii} of an m by n matrix A are called (*principal*) *diagonal elements*, and if $a_{ij} = 0$ for $i \neq j$, then A is called a *diagonal matrix*.

The diagonal elements as such may be defined as the elements which do not change their relative positions under transposition; e.g. a_{11} and a_{22} in the example of Df. 4.1.3.2.13 are unchanged by transposition.

Df. 4.1.3.2.17 A matrix $A = (a_{ij})$ is called a *diagonal matrix* if its elements off the diagonal are all zero, i.e. $a_{ij} = 0$ for $i \neq j$; it is called a *square matrix of order n* if it is an n by n matrix, having the same number of rows and columns, and a *scalar matrix* if it is square and furthermore $a_{ii} = c$, $c \in F$. In particular, if $a_{ii} = 1$, A is an *identity matrix*, denoted by $[1]$, and if $a_{ii} = 0$, it is then obviously a *zero matrix*, denoted by $[0]$, every element of which is zero.

Zero matrices are not limited to square matrices, since any m by n matrix may have all zero elements; moreover, the set \bar{M} of all m by n matrices contains zero-divisors, since the product of two non-zero matrices may be a zero matrix, e.g.,

$$A = \begin{bmatrix} a & b \\ 0 & 0 \end{bmatrix} \neq 0, \quad B = \begin{bmatrix} 0 & b \\ 0 & -a \end{bmatrix} \neq 0, \quad \text{but} \quad AB = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = 0$$

Hence cancellation cannot be carried out in \bar{M} .

Df. 4.1.3.2.17a The *determinant* $|A|$ of a square matrix $A = (a_{ij})$ of order n is a polynomial in a_{ij} of the form

$$|A| = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} = \sum_{j_1 j_2 \dots j_n} \epsilon_{j_1 j_2 \dots j_n} (a_{j_1 1} a_{j_2 2} \dots a_{j_n n})$$

where $\epsilon_{j_1 j_2 \dots j_n} = \begin{pmatrix} 1 & 2 & \dots & n \\ j_1 & j_2 & \dots & j_n \end{pmatrix} = p$ represents $n!$ permutations, j_1, j_2, \dots, j_n

being $1, 2, \dots, n$ in some order, yielding $p = 1$ if p is an even permutation and $p = -1$ if p is an odd permutation (cf. Df. 3.1.1.16).

Example:

If A is a matrix of order 2, then

$$|A| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = \epsilon_{12} a_{11} a_{22} + \epsilon_{21} a_{21} a_{12} = 1 \cdot a_{11} a_{22} + (-1) \cdot a_{21} a_{12} = a_{11} a_{22} - a_{21} a_{12}$$

The fundamental properties of determinants, with which the student is already familiar through College Algebra, are all deducible from this definition (cf. Prob. 23-29).

Note. $\epsilon_{j_1 j_2 \dots j_n} = \delta_{j_1 j_2 \dots j_n}^{1\ 2\ \dots\ n}$ is the generalized *Kronecker delta*.

Df. 4.1.3.2.18 The *cofactor* of the element a_{ij} of a square matrix $A = (a_{ij})$ is denoted by A_{ij} , which is stipulated by the following equation

$$|A| = \sum a_{ij} A_{ij}$$

which is summed either by i for fixed j or by j for fixed i (cf. Prob. 35).

Example:

If A is a square matrix of order 3, then

$$\begin{aligned} |A| &= |a_{ij}| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \\ &= \epsilon_{123} a_{11} a_{22} a_{33} + \epsilon_{132} a_{11} a_{32} a_{23} + \epsilon_{213} a_{21} a_{12} a_{33} + \epsilon_{231} a_{21} a_{32} a_{13} + \epsilon_{312} a_{31} a_{12} a_{23} + \epsilon_{321} a_{31} a_{22} a_{13} \\ &= a_{11} a_{22} a_{33} - a_{11} a_{32} a_{23} - a_{21} a_{12} a_{33} + a_{21} a_{32} a_{13} + a_{31} a_{12} a_{23} - a_{31} a_{22} a_{13} \\ &= a_{11}(a_{22} a_{33} - a_{32} a_{23}) + a_{21}(a_{32} a_{13} - a_{12} a_{33}) + a_{31}(a_{12} a_{23} - a_{22} a_{13}) \\ &= a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} + a_{21} \begin{vmatrix} a_{12} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} + a_{31} \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} \end{aligned}$$

where the cofactor A_{21} of a_{21} , for instance, is $(a_{32} a_{13} - a_{12} a_{33})$ or what is the same $-(a_{12} a_{33} + a_{32} a_{13})$, either of which may be given in a determinant form, as is quite obviously the case. Hence the cofactor is more conveniently defined in terms of determinants as follows:

Df. 4.1.3.2.18a If \bar{A}_{ij} denotes the $(n-1)$ by $(n-1)$ *submatrix*, called a *minor*, of an n by n matrix $A = (a_{ij})$ obtained by deleting the i -th row and the j -th column of A , then

$$A_{ij} = (-1)^{i+j} |\bar{A}_{ij}|$$

is called the *cofactor* of a_{ij} in A .

In the example given directly above, then,

$$\begin{aligned} |A| &= \sum_i a_{i1} A_{i1}, \quad i=1,2,3, \\ &= a_{11} A_{11} + a_{21} A_{21} + a_{31} A_{31} = (-1)^{1+1} a_{11} A_{11} + (-1)^{2+1} a_{21} A_{21} + (-1)^{3+1} a_{31} A_{31} \\ &= a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{21} \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{31} \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} \end{aligned}$$

which of course yields the same result as above. It must be emphasized, however, that the cofactors other than A_{11} , viz. $A_{12}, A_{13}, A_{21}, A_{22}, A_{23}, A_{31}, A_{32}, A_{33}$ yield, in this case, five other ways to evaluate $|A|$ (cf. Prob. 35 note).

Df. 4.1.3.2.19 The *adjoint* of a square matrix $A = (a_{ij})$ is also a square matrix, denoted by A^* , of the form $(A_{ij})^T$, where A_{ij} is the cofactor of a_{ij} .

In this matrix, then, each element is itself a determinant (cf. Prob. 37). A^* is thus found through two steps, first by finding the cofactors (in determinants with \pm signs) of all elements of A to form a matrix with the cofactors as its elements, then by transposing the matrix.

Df. 4.1.3.2.20 A square matrix A is said to be *nonsingular* if $|A| \neq 0$.

Another definition of nonsingularity, viz. in terms of the inverse of A , is also available, anticipating the following theorem.

Th. 4.1.3.2.21 If $|A| \neq 0$, where A is a matrix of order n , then there uniquely exists a square matrix B of the same order such that $AB = BA = [1]$, where $B = A^*/|A|$. (Cf. Prob. 39.)

After this theorem, then, A may be said to be nonsingular if there exists B such that $AB = BA = [1]$, i.e. if there exists A^{-1} for A .

Th. 4.1.3.2.22 The set \bar{A} of all square matrices of order n forms a noncommutative ring. (Cf. Prob. 42.)

The set \bar{A} is sometimes called a *total matrix algebra* (over a ring R or a field F) of order n , which is to satisfy Df. 4.1.3.2.1, and as such constitutes a *division algebra* (over R or F) if it satisfies, furthermore, Th. 4.1.3.2.21 in addition to Df. 4.1.3.2.1.

Solved Problems

1. Prove Th. 4.1.3.2.2 in terms of Df. 4.1.3.2.1.

PROOF:

Let $a, b \in F$ and $\alpha, \beta, \gamma \in V_n(F)$, i.e. $\alpha = (a_i)$, $\beta = (b_i)$, $\gamma = (c_i)$, $i = 1, 2, \dots, n$; then, by Df. 4.1.3.2.1,

$$\text{V1. } \alpha + \beta = (a_i) + (b_i) = (a_i + b_i) \in V_n(F)$$

$$\begin{aligned} \text{V2. } \alpha + (\beta + \gamma) &= (a_i) + ((b_i) + (c_i)) = (a_i) + (b_i + c_i) = (a_i + b_i + c_i) \\ &= ((a_i + b_i) + c_i) = (a_i + b_i) + \gamma = (\alpha + \beta) + \gamma \end{aligned}$$

$$\text{V3. } o = (O_i) \in V_n(F), \text{ called the zero vector (of order } n\text{).}$$

$$\text{V4. } \alpha^{-1} = (-a_i) \in V_n(F), \text{ since } \alpha + \alpha^{-1} = \alpha^{-1} + \alpha = (a_i - a_i) = (O_i) = o$$

$$\text{V5. } \alpha + \beta = (a_i) + (b_i) = (a_i + b_i) = (b_i + a_i) = \beta + \alpha$$

$$\text{V6. } a\alpha = a(a_i) = (aa_i) \in V_n(F)$$

$$\text{V7. (i) } a(\alpha + \beta) = a((a_i) + (b_i)) = a(a_i + b_i) = (a(a_i + b_i)) \in V_n(F)$$

$$\text{(ii) } (a + b)\alpha = (a + b)(a_i) = ((a + b)a_i) \in V_n(F)$$

$$\text{V8. (i) } (ab)\alpha = (ab)(a_i) = (aba_i) = (a(ba_i)) = a(ba_i) = a(b\alpha)$$

$$\text{(ii) } 1\alpha = 1(a_i) = (1 \cdot a_i) = (a_i) = \alpha.$$

2. If $a \in F$ and $\alpha \in V_n(F)$, then (i) $0 \cdot \alpha = o$, (ii) $(-a)\alpha = -a\alpha = a(-\alpha)$, (iii) $ao = o$.

PROOF:

(i) By V6, $0 \cdot \alpha = 0(a_i) = (0 \cdot a_i) = (O_i) = o$. (Note. 0 in $0 \cdot \alpha$ is a scalar, i.e. $0 \in F$, while o on the other side of the equation is a vector, i.e. $o \in V_n(F)$, viz. $o = (O_i) = (O_1, O_2, \dots, O_n)$, the zero vector, where each of O_i is 0; cf. Prob. 1, V3.)

(ii) By V6, $(-a)\alpha = (-a)(a_i) = (-aa_i) = ((-1)aa_i) = (-1)(aa_i) = (-1)a\alpha = -a\alpha$.

Furthermore, by V6, $-a\alpha = (-aa_i) = ((-1) \cdot aa_i) = (a(-1)a_i) = a(-a_i) = a(-\alpha)$.

(iii) As in (i), $ao = a(O_i) = (aO_i) = (O_i) = o$, by V6.3.

Second Proof. Since $\alpha = \alpha + o$ for every $\alpha \in V(F)$, by V3, it follows from V7, (i) that $a\alpha = a(\alpha + o) = a\alpha + ao$. Hence, adding $-a\alpha$ to both sides of the equation, $(-a\alpha) + a\alpha = (-a\alpha) + a\alpha + ao$, i.e. $o = o + ao$, viz. $o = ao$, by V3, completing the proof.

3. Prove that $\bar{R}(R)$, where R is the set of rational numbers and \bar{R} the set of real numbers, is a vector space.

PROOF:

Let $a, b \in R$ and $\alpha, \beta, \gamma \in \bar{R}$, then obviously $\alpha = (a_i)$, $\beta = (b_i)$, $\gamma = (c_i)$, where $i = 1$, and α, β, γ satisfy V1-5, since V1-5 \equiv F1-5 in this context. Furthermore, since the product of two rational numbers is again a rational number and the product of a rational number and a real number is a real number, $a, b, \alpha, \beta, \gamma$ satisfy also V6-8. Hence $\bar{R}(R)$ is a vector space, of order 1.

Note. It is similarly proved that $F(F)$ is also a vector space of order 1.

- *4. If L is the set of all functions which are solutions of the linear differential equation

$$y'' - 4y' + 3y = 0 \quad (1)$$

where y is a function of x , then L is a vector space.

PROOF:

If $a, b \in \bar{R}$ (as in Prob. 3) and $f_1(x), f_2(x), f_3(x) \in L$, satisfying the equation (1), then, from the theorems of the Calculus:

$$\text{V1. } f_1(x) + f_2(x) \in L, \text{ since } (f_1(x) + f_2(x))'' - 4(f_1(x) + f_2(x))' + 3(f_1(x) + f_2(x)) = (f_1''(x) - 4f_1'(x) + 3f_1(x)) + (f_2''(x) - 4f_2'(x) + 3f_2(x)) = 0$$

$$\text{V2. } f_1(x) + (f_2(x) + f_3(x)) = (f_1(x) + f_2(x)) + f_3(x), \text{ which follows from V1}$$

$$\text{V3. } 0 \in L, \text{ since } 0'' - 4 \cdot 0 + 3 \cdot 0 = 0$$

$$\text{V4. } (f_i(x))^{-1} = -f_i(x) \in L, i = 1, 2, \dots, \text{ since } f_i(x) + (-f_i(x)) = 0$$

$$\text{V5. } f_1(x) + f_2(x) = f_2(x) + f_1(x), \text{ which follows from V1}$$

$$\text{V6. } af_1(x) \in L, \text{ since } (af_1(x))'' - 4(af_1(x))' + 3(af_1(x)) = a(f_1''(x) - 4f_1'(x) + 3f_1(x)) = 0$$

V7-8 follow immediately from V6.

Hence L is a vector space over \bar{R} .

Note. In actual context, each of $f_i(x)$, $i = 1, 2, \dots$, is of the form $ae^x + be^{3x}$, $a, b \in \bar{R}$, since e^x and e^{3x} are linearly independent (cf. Df. 4.1.3.2.4).

5. Prove Th. 4.1.3.2.3.

PROOF:

If U is a subspace of $V(F)$, then $a\alpha + b\beta \in U$ by V2, 6.

Conversely, if $a\alpha + b\beta \in U$ for every $a, b \in F$ and $\alpha, \beta \in U$, then $\alpha + \beta \in U$, since $1\alpha + 1\beta = \alpha + \beta$, and also $0 \in U$, since $a\alpha + 0\beta = a\alpha \in U$. The other properties of V1-8 are consequently satisfied. Hence U is a subspace of $V(F)$.

Note the similarity between this theorem and Th. 4.1.1.7 or, further back, Th. 3.2.1.2. Note, also, that $a\alpha + b\beta$ embodies *linearity* at one stroke, combining the effects of V6-8.

6. Prove, or disprove, that the following sets of real functions defined on $0 \leq x \leq 2$ are subspaces of the vector space of all such functions:

(i) F_1 : all functions f such that $f(1) = 3f(2)$,

(ii) F_2 : all functions f such that $f(x) = f(x-1)$ for all x ,

(iii) F_3 : all polynomials of degree 5.

PROOF:

By Th. 4.1.3.2.3,

(i) F_1 forms a subspace, since $f_1, f_2 \in F_1$ implies $f_1(1) = 3f_1(2)$ and $f_2(1) = 3f_2(2)$, which in turn implies $f_3 \in F_1$ such that $f_3 = f_1(1) + f_2(x) = 3f_1(2) + 3f_2(2) = 3(f_1(2) + f_2(2)) = 3f_3(2)$.

- (ii) Likewise, $f_1, f_2 \in F_2$ implies $f_3 \in F_2$ such that $f_3(x) = f_1(x) + f_2(x) = f_1(x-1) + f_2(x-1) = f_3(x-1)$, and F_2 does form a subspace.
- (iii) F_3 does not form a subspace, since $f_1, f_2 \in F_3$, i.e. $f_1 = \sum_i a_i x^i$ and $f_2 = \sum_i b_i x^i$, $i=0,1,\dots,5$, does not always imply $f_3 \in F_3$ such that $f_3 = f_1 + f_2 = \sum_i c_i x^i$, $i=0,1,\dots,5$; e.g. if $f_1 = -f_2$, then $f_3 = f_1 + f_2 = 0$, which is definitely not a polynomial of degree 5. (Note. F'_3 : all polynomials of degree less than 5, including 0, however, is a subspace, as can be readily verified.)

7. Prove Th. 4.1.3.2.5.

PROOF:

By Df. 4.1.3.2.4, the elements of S are of the form $\sum_i a_i \alpha_i$, $\sum_i b_i \alpha_i$, etc., where $i=1,2,\dots,k$ for $1 \leq k \leq n$. Hence, for any $r, s, t \in F$,

$$r\left(\sum_i a_i \alpha_i\right) + s\left(\sum_i b_i \alpha_i\right) = \sum_i r a_i \alpha_i + \sum_i s b_i \alpha_i = \sum_i (r a_i + s b_i) \alpha_i = \sum_i t_i \alpha_i \in S$$

where $t_i = r a_i + s b_i$, $i=1,2,\dots,k$. Hence, by Df. 4.1.3.2.4, S is a subspace of $V_n(F)$. (Cf. Prob. 5 note.)

8. The meet of any two subspaces, say U and W , of a vector space V is again a subspace of V ; so is their linear sum, defined by $\{\alpha + \beta\}$, for any $\alpha \in U$ and $\beta \in W$.

PROOF:

- (i) Since the meet $M = U \cap W$ contains all vectors which belong to both U and W , $\alpha, \beta \in M$ implies that $\alpha + \beta \in M$, $a\alpha \in M$, $b\beta \in M$; in short, $a\alpha + b\beta \in M$. Hence, by Df. 4.1.3.2.4, M is a subspace of V .
- (ii) The linear sum, by definition itself, at once satisfies Df. 4.1.3.2.4, proving itself to be a subspace of V .

9. Prove Th. 4.1.3.2.6.

PROOF:

If $\alpha_1, \alpha_2, \dots, \alpha_k$ are linearly dependent, then, by Df. 4.1.3.2.3,

$$\alpha = a_1 \alpha_1 + a_2 \alpha_2 + \dots + a_k \alpha_k = 0 \quad (1)$$

where at least one of a_i , $i=1,2,\dots,k$, is not zero. Hence, for any $a_i \neq 0$, (1) is changed to

$$a_i \alpha_i = -(a_1 \alpha_1 + a_2 \alpha_2 + \dots + a_{i-1} \alpha_{i-1} + a_{i+1} \alpha_{i+1} + \dots + a_k \alpha_k)$$

i.e.,

$$\alpha_i = b_1 \alpha_1 + b_2 \alpha_2 + \dots + b_{i-1} \alpha_{i-1} + b_{i+1} \alpha_{i+1} + \dots + b_n \alpha_n$$

where $b_j = -a_i^{-1} a_j$, for fixed i and $j = 1, 2, \dots, i-1, i+1, \dots, n$.

Conversely, if α_i is spanned by other vectors than itself, i.e.,

$$\alpha_i = c_1 \alpha_1 + c_2 \alpha_2 + \dots + c_{i-1} \alpha_{i-1} + c_{i+1} \alpha_{i+1} + \dots + c_k \alpha_k$$

then

$$c_1 \alpha_1 + c_2 \alpha_2 + \dots + c_{i-1} \alpha_{i-1} + (-1)\alpha_i + c_{i+1} \alpha_{i+1} + \dots + c_k \alpha_k = 0$$

proving that the vectors α_i , $i=1,2,\dots,k$, are linearly dependent, which completes the proof.

10. The basis of a finite-dimensional vector space is an invariant.

PROOF:

Let the finite-dimensional vector space be $V_n(F)$, and assume that $V_n(F)$ has two bases of m and n , say $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_m)$, $\beta = (\beta_1, \beta_2, \dots, \beta_n)$, and $\alpha, \beta \in V_n(F)$. Then $m \geq n$, since α spans $V_n(F)$, and β , consisting of base vectors, is linearly independent. Likewise $n \geq m$, since β also spans $V_n(F)$, and α is linearly independent. Hence $m = n$, completing the proof.

11. Addition is associative and commutative among matrices of the same order.

PROOF:

Let A, B, C be of the same size, say $m \times n$, i.e. $A = (a_{ij})$, $B = (b_{ij})$, $C = (c_{ij})$, where $i = 1, 2, \dots, m$, $j = 1, 2, \dots, n$; then, by Df. 4.1.3.2.10,

$$\begin{aligned} A + (B + C) &= (a_{ij}) + ((b_{ij}) + (c_{ij})) = (a_{ij}) + (b_{ij} + c_{ij}) = (a_{ij} + b_{ij} + c_{ij}) \\ &= ((a_{ij} + b_{ij}) + c_{ij}) = (a_{ij} + b_{ij}) + (c_{ij}) = (A + B) + C \end{aligned}$$

$$\text{and} \quad A + B = (a_{ij}) + (b_{ij}) = (a_{ij} + b_{ij}) = (b_{ij} + a_{ij}) = (b_{ij}) + (a_{ij}) = B + A$$

completing the proof.

Note that the proof has been carried out, as in Prob. 1 above, on the strength of the additive associativity and commutativity of a_{ij}, b_{ij}, c_{ij} , which are any three elements of a field F , where additive associativity (F2) and commutativity (F5) evidently hold.

12. Prove Th. 4.1.3.2.11.

PROOF:

Let $a, b \in F$ and $A, B, C \in \bar{M}$, i.e. $A = (a_{ij})$, $B = (b_{ij})$, $C = (c_{ij})$, where $i = 1, 2, \dots, m$, $j = 1, 2, \dots, n$; then V1 and V6 are already given by Df. 4.1.3.2.1 itself, and also V2 and V5 by Prob. 11. Furthermore:

V3. Additive identity is the m by n zero matrix, O , viz.

$$O = (O_{ij}) = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}$$

V4. Additive inverse $-A = (-a_{ij})$ for every $A \in \bar{M}$, since $A + (-A) = (-A) + A = O$.

V7. $a(A + B) = a((a_{ij}) + (b_{ij})) = a(a_{ij}) + a(b_{ij}) = aA + aB$

$$(a + b)A = (a + b)(a_{ij}) = a(a_{ij}) + b(a_{ij}) = aA + bA$$

V8. $(ab)A = (ab)(a_{ij}) = (aba_{ij}) = (a(ba_{ij})) = a(b(a_{ij})) = a(bA)$

$$1A = 1(a_{ij}) = (1 \cdot a_{ij}) = (a_{ij}) = A.$$

Hence \bar{M} , satisfying all of V1-8, is a vector space over F .

13. Given $A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ and $B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, verify $AB \neq BA$.

PROOF:

By Df. 4.1.3.2.12,

$$AB = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 \cdot 0 + 0 \cdot 1 & 1 \cdot 1 + 0 \cdot 0 \\ 0 \cdot 0 - 1 \cdot 1 & 0 \cdot 1 - 1 \cdot 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

$$\text{and} \quad BA = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 0 \cdot 1 + 1 \cdot 0 & 0 \cdot 0 - 1 \cdot 1 \\ 1 \cdot 1 + 0 \cdot 0 & 1 \cdot 0 - 0 \cdot 1 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

Hence, by Df. 4.1.3.2.9, $AB \neq BA$.

Note. This example alone is already enough to establish the non-commutativity in \bar{M} , since a single counterexample is sufficient to disprove a theorem.

14. Given $A = \begin{bmatrix} 1 & 2 & 1 \\ 3 & 4 & 2 \\ 1 & 3 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 10 & -4 & -1 \\ -11 & 5 & 0 \\ 9 & -5 & 1 \end{bmatrix}$, verify $AB = BA$.

PROOF:

$$\text{As in Prob. 11 above, by Df. 4.1.3.2.11,} \quad AB = \begin{bmatrix} -3 & 1 & 0 \\ 4 & -2 & -1 \\ -5 & 1 & 1 \end{bmatrix} = BA.$$

Note. Commutativity, therefore, holds *sometimes* in \bar{M} ; it may even hold on *all* occasions with respect to *some* matrices (cf. Prob. 38 below), although it cannot be said about *all* matrices, as was proved by Prob. 13.

15. Given $A = \begin{bmatrix} a & a \\ b & b \end{bmatrix}$, prove that there exists no matrix X such that $AX = E$,

where $E = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = [1]$.

PROOF:

Assume that there exists X , i.e. $X = \begin{bmatrix} c & d \\ e & f \end{bmatrix}$, such that

$$AX = \begin{bmatrix} a & a \\ b & b \end{bmatrix} \begin{bmatrix} c & d \\ e & f \end{bmatrix} = \begin{bmatrix} a(c+e) & a(d+f) \\ b(c+e) & b(d+f) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = E$$

Then, by Df. 4.1.3.2.9, it must follow that

$$\begin{array}{ll} \text{(i)} & a(c+e) = 1 \\ \text{(ii)} & a(d+f) = 0 \\ \text{(iii)} & b(c+e) = 0 \\ \text{(iv)} & b(d+f) = 1 \end{array}$$

hold simultaneously. This leads to a contradiction, since (i) implies $a \neq 0$, which implies, by (ii), $d+f = 0$, which in turn implies $b(d+f) = 0$, contradictory to (iv). Hence there exists no X such that $AX = E$.

Note. As will be seen below (cf. Prob. 38-39), the non-existence of such a matrix X is simply due to the singularity (cf. Df. 4.1.3.2.20) of A , whose determinant is obviously 0 (cf. Prob. 25).

16. Given $A = \begin{bmatrix} a & ac \\ b & bc \end{bmatrix}$ and $B = \begin{bmatrix} cd & ce \\ -d & -e \end{bmatrix}$, prove that they are zero-divisors (i.e. *nilfactors*), viz. $AB = 0$ when $A \neq 0$ and $B \neq 0$.

PROOF:

$$AB = \begin{bmatrix} a \cdot cd - ac \cdot d & a \cdot ce - ac \cdot e \\ b \cdot cd - bc \cdot d & b \cdot ce - bc \cdot e \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = 0$$

Note. $AB \neq BA$ in this context and also $BA \neq 0$, as can be readily verified.

17. Matrix multiplication is associative, if conformability is assured; i.e. $A(BC) = (AB)C$ if A, B, C are p by q , q by r , r by s matrices respectively.

PROOF:

Let $A = (a_{hi})$, $B = (b_{ij})$, $C = (c_{jk})$, where $h = 1, 2, \dots, p$, $i = 1, 2, \dots, q$, $j = 1, 2, \dots, r$, $k = 1, 2, \dots, s$; then

$$\begin{aligned} A(BC) &= (a_{hi})((b_{ij})(c_{jk})) = (a_{hi})\left(\sum_j b_{ij} c_{jk}\right) = \sum_i a_{hi} \left(\sum_j b_{ij} c_{jk}\right) = \sum_i \sum_j a_{hi} b_{ij} c_{jk} \\ &= \sum_j \left(\sum_i a_{hi} b_{ij}\right) c_{jk} = \left(\sum_i a_{hi} b_{ij}\right)(c_{jk}) = ((a_{hi})(b_{ij}))(c_{jk}) = (AB)C \end{aligned}$$

Note. Square matrices are always associative under multiplication, since conformability is always assured for them. The conformability alone, however, does not always guarantee associativity, since infinite matrices are always conformable to each other, but it is not always the case that $A(BC) = (AB)C$; e.g.,

$$\begin{aligned} A(BC) &= \begin{bmatrix} 1 & 1 & 1 & 1 & \dots \\ 0 & 1 & 1 & 1 & \dots \\ 0 & 0 & 1 & 1 & \dots \\ 0 & 0 & 0 & 1 & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix} \left(\begin{bmatrix} 1 & -1 & 0 & 0 & \dots \\ 0 & 1 & -1 & 0 & \dots \\ 0 & 0 & 1 & -1 & \dots \\ 0 & 0 & 0 & 1 & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & \dots \\ 1 & 1 & 1 & 1 & \dots \\ 1 & 1 & 1 & 1 & \dots \\ 1 & 1 & 1 & 1 & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix} \right) \\ &= \begin{bmatrix} 1 & 1 & 1 & 1 & \dots \\ 1 & 1 & 1 & 1 & \dots \\ 1 & 1 & 1 & 1 & \dots \\ 1 & 1 & 1 & 1 & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix} \neq \begin{bmatrix} 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix} = (AB)C \end{aligned}$$

18. If $AB = BA$, then $(AB)^n = A^n B^n$.

PROOF:

If $n = 1$, then $(AB)^1 = A^1 B^1$, since it is the given hypothesis itself.

If $n = 2$, then, by the associative law proved in Prob. 15 (which holds here because of the conformability assured by the given commutativity),

$$(AB)^2 = (AB)(AB) = (ABA)B = (AAB)B = (A^2 B)B = A^2(BB) = A^2 B^2$$

In general, if $(AB)^k = A^k B^k$, then

$$\begin{aligned} (AB)^{k+1} &= (AB)^k(AB) = A^k B^k AB = A^k (B^k A)B = A^k (B^{k-1} BA)B = A^k (B^{k-1} AB)B \\ &= A^k (B^{k-2} BAB)B = A^k (B^{k-2} ABB)B = \dots = A^k AB^k B = A^{k+1} B^{k+1} \end{aligned}$$

Hence, in general, $AB = BA$ implies $(AB)^n = A^n B^n$.

19. Matrix multiplication is distributive under addition, if conformability is assured; i.e. $A(B+C) = AB+AC$ if A is a p by q matrix and B and C are q by r matrices; also $(A+B)C = AC+BC$ if A and B are p by q matrices and C is a q by r matrix.

PROOF:

(i) By Df. 4.1.3.2.10, 12, $A = (a_{ij})$, $B = (b_{jk})$, $C = (c_{jk})$, where $i = 1, 2, \dots, p$, $j = 1, 2, \dots, q$, $k = 1, 2, \dots, r$, implies

$$A(B+C) = (a_{ij})((b_{jk}) + (c_{jk})) = (a_{ij})(b_{jk} + c_{jk}) = \sum_j a_{ij} (b_{jk} + c_{jk}) = \sum_j a_{ij} b_{jk} + \sum_j a_{ij} c_{jk} = AB + AC$$

(ii) Likewise $(A+B)C = AC+BC$.

Note. Distributive laws hold unconditionally for square matrices, since conformability is always assured for them.

20. If A and B are square matrices of the same order, find $(A+B)^2$ and $(A+B)^3$.

Solution:

By Df. 4.1.3.2.12 and Prob. 18,

$$(A+B)^2 = (A+B)(A+B) = (A+B)A + (A+B)B = A^2 + BA + AB + B^2$$

and

$$\begin{aligned} (A+B)^3 &= (A+B)^2(A+B) = (A^2 + BA + AB + B^2)(A+B) \\ &= A^3 + BA^2 + ABA + B^2A + A^2B + BAB + AB^2 + B^3 \end{aligned}$$

Note. The binomial coefficients (or Pascal's triangle) cannot be introduced here, because $AB \neq BA$ in general.

21. If A and B are two m by n matrices over F , then $(A^T)^T = A$ and $(aA+bB)^T = aA^T + bB^T$ for every $a, b \in F$.

PROOF:

(i) $(A^T)^T = ((a_{ij})^T)^T = (a_{ji})^T = (a_{ij}) = A$, where $i = 1, 2, \dots, m$, $j = 1, 2, \dots, n$.

(ii) $(aA+bB)^T = (a(a_{ij}) + b(b_{ij}))^T = (aa_{ij} + bb_{ij})^T = (aa_{ji} + bb_{ji})$
 $= (aa_{ji}) + (bb_{ji}) = a(a_{ji}) + b(b_{ji}) = a(a_{ij})^T + b(b_{ij})^T = aA^T + bB^T$

where i and j are summed as above.

22. Prove Th. 4.1.3.2.15.

PROOF:

Since A and B are conformable in both directions, let A be an m by n matrix and B an n by m matrix, i.e. $A = (a_{ij})$, $B = (b_{ji})$, where $i = 1, 2, \dots, m$, $j = 1, 2, \dots, n$. Then, by Df. 4.1.3.2.12, 14,

$$(AB)^T = ((a_{ij})(b_{ji}))^T = \left(\sum_j a_{ij} b_{ji}\right)^T = \left(\sum_j b_{ij} a_{ji}\right)^T = \left(\sum_j b_{ij} a_{ji}\right) = (b_{ij})(a_{ji}) = B^T A^T$$

where $\sum_j a_{ij} b_{ji} = \sum_j b_{ij} a_{ji}$ because of commutativity in F , and also $\left(\sum_j b_{ij} a_{ji}\right)^T = \left(\sum_j b_{ij} a_{ji}\right)$, because of the specific conformability assumed at the start, due to which the interchange of i at both ends of $\sum_j b_{ij} a_{ji}$ brings forth $\sum_j b_{ij} a_{ji}$ itself, completing the proof.

23. If A is a square matrix, then $|A^T| = |A|$.

PROOF:

Let A be of order n , i.e. $A = (a_{ij})$, $i, j = 1, 2, \dots, n$; then $A^T = (a_{ji})$ and, by Df. 4.1.3.2.17,

$$\begin{aligned} |A^T| &= \sum_{j=1}^n \epsilon_{j_1 j_2 \dots j_n} (a_{j_1 1} a_{j_2 2} \dots a_{j_n n}) = \sum_{j=1}^n \epsilon_{j_1 j_2 \dots j_n} (a_{1 j_1} a_{2 j_2} \dots a_{n j_n}) \\ &= \sum_{i=1}^n \epsilon^{i_1 i_2 \dots i_n} (a_{i_1 1} a_{i_2 2} \dots a_{i_n n}) = |A| \end{aligned}$$

since the summation by j has exactly the same effect as the summation by i , both going from 1 to n , and the interchange of i and j does not affect the transposition, i.e. $\epsilon^{i_1 i_2 \dots i_n} = \epsilon_{j_1 j_2 \dots j_n}$, which is quite obvious from the definition of permutation.

Note. The only difference, a superficial one, is that the first (by i) is summed *row-wise* and the second (by j) *column-wise*, which cannot affect the final result as long as A is a square matrix; e.g.,

$$|A^T| = \begin{vmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{vmatrix} = \epsilon^{12} a_{11} a_{22} + \epsilon^{21} a_{12} a_{21} = \epsilon^{12} a_{11} a_{22} + \epsilon^{21} a_{21} a_{12} = |A|$$

Hence a determinant may be summed over either row or column subscripts, yielding the same result.

24. If B is a matrix obtained by interchanging two rows (or columns) of a square matrix A , then $|B| = -|A|$.

PROOF:

Let A be a matrix of order n , i.e. $A = (a_{ij})$, $i, j = 1, 2, \dots, n$, and interchange two rows, say i_k -th row and i_{k+r} -th row, where $k = 1, 2, \dots, n$ and $r = 0, 1, \dots, n-1$; then by hypothesis and Df. 4.1.3.2.17,

$$\begin{aligned} |B| &= \sum_{i=1}^n \epsilon^{i_1 i_2 \dots i_{k+r} \dots i_k \dots i_n} (a_{i_1 1} a_{i_2 2} \dots a_{i_{k+r} k+r} \dots a_{i_k k} \dots a_{i_n n}) \\ &= (-1) \sum_{i=1}^n \epsilon^{i_1 i_2 \dots i_k \dots i_{k+r} \dots i_{k+r} \dots i_n} (a_{i_1 1} a_{i_2 2} \dots a_{i_k k} \dots a_{i_{k+r} k+r} \dots a_{i_n n}) = (-1) |A| = -|A| \end{aligned}$$

25. If two rows (or columns) of a square matrix A are identical, then $|A| = 0$.

PROOF:

If B is a matrix obtained by interchanging two identical rows (or columns) of A , then $B = A$, which implies $|B| = |A|$, since the interchange of the identical rows (or columns) yields the same matrix A . On the other hand it follows, by Prob. 24 above, that the interchange of two rows (or columns) yields, *ipso facto*,

$$|B| = -|A|$$

where $|B| = |A|$ in this context. Hence $|A| = -|A|$, i.e. $2|A| = 0$, which implies $|A| = 0$, completing the proof.

26. If B is a matrix obtained by multiplying every element of one row (or column) of a square matrix A by a constant c , then $|B| = c|A|$.

PROOF:

Let the i_k -th row be multiplied by c ; then, by hypothesis,

$$\begin{aligned} |B| &= \sum_{i=1}^n \epsilon^{i_1 i_2 \dots i_k \dots i_n} (b_{i_1 1} b_{i_2 2} \dots b_{i_k k} \dots b_{i_n n}) = \sum_{i=1}^n \epsilon^{i_1 i_2 \dots i_k \dots i_n} (a_{i_1 1} a_{i_2 2} \dots c a_{i_k k} \dots a_{i_n n}) \\ &= \sum_{i=1}^n \epsilon^{i_1 i_2 \dots i_k \dots i_n} c (a_{i_1 1} a_{i_2 2} \dots a_{i_k k} \dots a_{i_n n}) \\ &= c \left(\sum_{i=1}^n \epsilon^{i_1 i_2 \dots i_k \dots i_n} (a_{i_1 1} a_{i_2 2} \dots a_{i_k k} \dots a_{i_n n}) \right) = c|A| \end{aligned}$$

27. If A is a square matrix of order n , then $|cA| = c^n |A|$.

PROOF:

By Prob. 26 (or directly by Df. 4.1.3.2.17),

$$\begin{aligned} |cA| &= |c(a_{ij})| = |c(a_{ij})| \\ &= \sum_{i=1}^n \epsilon^{i_1 i_2 \dots i_n} (c a_{i_1 1} c a_{i_2 2} \dots c a_{i_n n}) = \sum_{i=1}^n \epsilon^{i_1 i_2 \dots i_n} c^n (a_{i_1 1} a_{i_2 2} \dots a_{i_n n}) = c^n |A| \end{aligned}$$

28. If B is a matrix obtained by adding to each element of a row (or column) of a square matrix A a constant multiple of the corresponding element of another row (or column), then $|B| = |A|$.

PROOF:

By hypothesis and by Prob. 25,

$$\begin{aligned}
 |B| &= \sum_{i=1}^n \epsilon^{i_1 i_2 \dots i_r \dots i_s \dots i_n} (b_{i_1 1} b_{i_2 2} \dots b_{i_r r} \dots b_{i_s s} \dots b_{i_n n}) \\
 &= \sum_{i=1}^n \epsilon^{i_1 i_2 \dots i_r \dots i_s \dots i_n} (a_{i_1 1} a_{i_2 2} \dots (a_{i_r r} + c a_{i_s s}) \dots a_{i_s s} \dots a_{i_n n}) \\
 &= \sum_{i=1}^n \epsilon^{i_1 i_2 \dots i_r \dots i_s \dots i_n} (a_{i_1 1} a_{i_2 2} \dots a_{i_r r} \dots a_{i_s s} \dots a_{i_n n}) \\
 &\quad + \sum_{i=1}^n \epsilon^{i_1 i_2 \dots i_r \dots i_s \dots i_n} (a_{i_1 1} a_{i_2 2} \dots c a_{i_s s} \dots a_{i_s s} \dots a_{i_n n}) \\
 &= |A| + c \left(\sum_{i=1}^n \epsilon^{i_1 i_2 \dots i_r \dots i_s \dots i_n} (a_{i_1 1} a_{i_2 2} \dots a_{i_s s} \dots a_{i_s s} \dots a_{i_n n}) \right) \\
 &= |A| + c \cdot 0 = |A|
 \end{aligned}$$

Note. For example:

$$\begin{aligned}
 |B| &= \begin{vmatrix} a & ka + b \\ c & kc + d \end{vmatrix} = \begin{vmatrix} a & ka \\ c & kc \end{vmatrix} + \begin{vmatrix} a & b \\ c & d \end{vmatrix} = k \begin{vmatrix} a & a \\ c & c \end{vmatrix} + \begin{vmatrix} a & b \\ c & d \end{vmatrix} \\
 &= k \cdot 0 + \begin{vmatrix} a & b \\ c & d \end{vmatrix} = |A|
 \end{aligned}$$

29. If A, B, C are three square matrices of the same order and are identical except for their respective k -th rows (or columns), where the k -th row (or column) of C is the matrix sum of the k -th rows (or columns) of A and B , then $|C| = |A| + |B|$.

PROOF:

Let the order of A, B, C be n ; then, by hypothesis and Prob. 28,

$$\begin{aligned}
 |C| &= \sum_{i=1}^n \epsilon^{i_1 i_2 \dots i_k \dots i_n} (c_{i_1 1} c_{i_2 2} \dots c_{i_k k} \dots c_{i_n n}) \\
 &= \sum_{i=1}^n \epsilon^{i_1 i_2 \dots i_k \dots i_n} (c_{i_1 1} c_{i_2 2} \dots (a_{i_k k} + b_{i_k k}) \dots c_{i_n n}) \\
 &= \sum_{i=1}^n \epsilon^{i_1 i_2 \dots i_k \dots i_n} (c_{i_1 1} c_{i_2 2} \dots a_{i_k k} \dots c_{i_n n}) + \sum_{i=1}^n \epsilon^{i_1 i_2 \dots i_k \dots i_n} (c_{i_1 1} c_{i_2 2} \dots b_{i_k k} \dots c_{i_n n}) \\
 &= \sum_{i=1}^n \epsilon^{i_1 i_2 \dots i_k \dots i_n} (a_{i_1 1} a_{i_2 2} \dots a_{i_k k} \dots a_{i_n n}) + \sum_{i=1}^n \epsilon^{i_1 i_2 \dots i_k \dots i_n} (b_{i_1 1} b_{i_2 2} \dots b_{i_k k} \dots b_{i_n n}) \\
 &= |A| + |B|
 \end{aligned}$$

Note. For example:

$$\begin{vmatrix} a + a' & b + b' \\ c & d \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \begin{vmatrix} a' & b' \\ c & d \end{vmatrix}$$

30. If $D \neq 0$, where

$$D = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}$$

then n vectors $\alpha_1 = (a_{11}, a_{12}, \dots, a_{1n})$, $\alpha_2 = (a_{21}, a_{22}, \dots, a_{2n})$, \dots , $\alpha_n = (a_{n1}, a_{n2}, \dots, a_{nn})$ are linearly independent.

PROOF:

If $D \neq 0$, then the following n simultaneous linear equations

$$a_{k1} x_1 + a_{k2} x_2 + \dots + a_{kn} x_n = a_k, \quad k = 1, 2, \dots, n$$

are found, by the so-called *Cramer's rule* (cf. College Algebra), to have n solutions of the form:

since the non-zero terms of the total summation are only those for which the set of k_1, k_2, \dots, k_n is some ordering of $1, 2, \dots, n$. Hence

$$|C| = \sum_{k=1}^n \epsilon^{k_1 k_2 \dots k_n} |A| b_{k_1 1} b_{k_2 2} \dots b_{k_n n} = |A| \sum_{k=1}^n \epsilon^{k_1 k_2 \dots k_n} b_{k_1 1} b_{k_2 2} \dots b_{k_n n} = |A| |B|$$

34. Given $x^3 = 1$ and $x^2 + x + 1 = 0$ (i.e. x is the cubic root of 1), evaluate D^2 , where

$$D = \begin{vmatrix} 1 & x & x^2 & x^3 \\ x & x^2 & x^3 & 1 \\ x^2 & x^3 & 1 & x \\ x^3 & 1 & x & x^2 \end{vmatrix}$$

Solution:

By Prob. 33,

$$\begin{aligned} D^2 &= D \cdot D = \begin{vmatrix} 1+x+x^2+1 & x+1+x^2+1 & x^2+x+x^2+x & 1+x+1+x^2 \\ x+1+x^2+1 & x^2+x+1+1 & 1+x^2+1+x & x+x^2+x+x^2 \\ x^2+x+x^2+x & 1+x^2+1+x & x+1+1+x^2 & x^2+1+x+1 \\ 1+x+1+x^2 & x+x^2+x+x^2 & x^2+1+x+1 & 1+1+x^2+x \end{vmatrix} \\ &= \begin{vmatrix} 1 & 1 & -2 & 1 \\ 1 & 1 & 1 & -2 \\ -2 & 1 & 1 & 1 \\ 1 & -2 & 1 & 1 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 3 & -3 \\ -2 & 3 & -3 & 3 \\ 1 & -3 & 3 & 0 \end{vmatrix} = 27 \begin{vmatrix} 0 & 1 & -1 \\ 1 & -1 & 0 \\ -1 & 1 & 0 \end{vmatrix} \\ &= -27, \text{ i.e. } D = \pm 3\sqrt{3}i \end{aligned}$$

35. If A_{ij} is the cofactor of a_{ij} of a square matrix $A = (a_{ij})$ of order n , then

$$\sum_{i=1}^n a_{ij} A_{ik} = |A| \delta_{jk} \quad \text{and} \quad \sum_{j=1}^n a_{ij} A_{kj} = |A| \delta_{ik}$$

where $i, j, k = 1, 2, \dots, n$, and δ_{ij} is the *Kronecker delta*, viz. $\delta_{ij} = 1$ or 0 according as $i = j$ or $i \neq j$ respectively.

PROOF:

It follows directly from Df. 4.1.3.2.18 and the definition of the Kronecker delta that if $j \neq k$, then

$$\sum_{i=1}^n a_{ij} A_{ik} = |A| \delta_{jk} = 0$$

and if $j = k$, then

$$\sum_{i=1}^n a_{ij} A_{ik} = \sum_{i=1}^n a_{ij} A_{ij} = |A| \delta_{jj} = |A| \cdot 1 = |A|$$

which is equivalent to

$$\sum a_{ij} A_{ij} = |A|, \quad i, j = 1, 2, \dots, n$$

i.e. Df. 4.1.3.2.18 itself, verifying the validity of the first formula.

The second formula can be validated likewise, completing the proof.

Note. For example, $A = (a_{ij})$, $i, j = 1, 2, 3$, implies

$$\begin{aligned} |A| &= \sum_i a_{i1} A_{i1} = a_{11} A_{11} + a_{21} A_{21} + a_{31} A_{31} \\ &= \sum_i a_{i2} A_{i2} = a_{12} A_{12} + a_{22} A_{22} + a_{32} A_{32} \\ &= \sum_i a_{i3} A_{i3} = a_{13} A_{13} + a_{23} A_{23} + a_{33} A_{33} \\ &= \sum_j a_{1j} A_{1j} = a_{11} A_{11} + a_{12} A_{12} + a_{13} A_{13} \\ &= \sum_j a_{2j} A_{2j} = a_{21} A_{21} + a_{22} A_{22} + a_{23} A_{23} \\ &= \sum_j a_{3j} A_{3j} = a_{31} A_{31} + a_{32} A_{32} + a_{33} A_{33} \\ &\neq \sum_i a_{i1} A_{i3} = a_{11} A_{13} + a_{21} A_{23} + a_{31} A_{33}, \text{ etc.,} \\ &\neq \sum_j a_{2j} A_{3j} = a_{21} A_{31} + a_{22} A_{32} + a_{23} A_{33}, \text{ etc.} \end{aligned}$$

36. If $D = |a_{ij}|$, then $D^{n-1} = |A_{ij}|$, $i, j = 1, 2, \dots, n$.

PROOF:

Since $D = |a_{ij}| = |a_{ji}|$, by Prob. 25, it follows from Prob. 33 and Prob. 35 that

$$\begin{aligned}
 |a_{ji}| |A_{ij}| &= \begin{vmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \dots & \dots & \dots & \dots \\ a_{1n} & a_{2n} & \dots & a_{nn} \end{vmatrix} \begin{vmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \dots & \dots & \dots & \dots \\ A_{n1} & A_{n2} & \dots & A_{nn} \end{vmatrix} \\
 &= \begin{vmatrix} a_{11}A_{11} + a_{21}A_{21} + \dots + a_{n1}A_{n1} & a_{11}A_{12} + a_{21}A_{22} + \dots + a_{n1}A_{n2} & \dots & a_{11}A_{1n} + a_{21}A_{2n} + \dots + a_{n1}A_{nn} \\ a_{12}A_{11} + a_{22}A_{21} + \dots + a_{n2}A_{n1} & a_{12}A_{12} + a_{22}A_{22} + \dots + a_{n2}A_{n2} & \dots & a_{12}A_{1n} + a_{22}A_{2n} + \dots + a_{n2}A_{nn} \\ \dots & \dots & \dots & \dots \\ a_{1n}A_{11} + a_{2n}A_{21} + \dots + a_{nn}A_{n1} & a_{1n}A_{12} + a_{2n}A_{22} + \dots + a_{nn}A_{n2} & \dots & a_{1n}A_{1n} + a_{2n}A_{2n} + \dots + a_{nn}A_{nn} \end{vmatrix} \\
 &= \begin{vmatrix} \sum_{i=1}^n a_{i1}A_{i1} & \sum_{i=1}^n a_{i1}A_{i2} & \dots & \sum_{i=1}^n a_{i1}A_{in} \\ \sum_{i=1}^n a_{i2}A_{i1} & \sum_{i=1}^n a_{i2}A_{i2} & \dots & \sum_{i=1}^n a_{i2}A_{in} \\ \dots & \dots & \dots & \dots \\ \sum_{i=1}^n a_{in}A_{i1} & \sum_{i=1}^n a_{in}A_{i2} & \dots & \sum_{i=1}^n a_{in}A_{in} \end{vmatrix} = \begin{vmatrix} D & 0 & \dots & 0 \\ 0 & D & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & D \end{vmatrix} = D^n
 \end{aligned}$$

Hence $D|A_{ij}| = D^n$, i.e. $|A_{ij}| = D^{n-1}$.

37. Find the adjoint A^* of $A = (a_{ij})$, $i, j = 1, 2, 3$.

Solution:

By Df. 4.1.3.2.18-19,

$$\begin{aligned}
 A^* &= (A_{ij})' = \begin{vmatrix} A_{11} & A_{21} & A_{31} \\ A_{12} & A_{22} & A_{32} \\ A_{13} & A_{23} & A_{33} \end{vmatrix} = \begin{vmatrix} (-1)^{1+1} \bar{A}_{11} & (-1)^{2+1} \bar{A}_{21} & (-1)^{3+1} \bar{A}_{31} \\ (-1)^{1+2} \bar{A}_{12} & (-1)^{2+2} \bar{A}_{22} & (-1)^{3+2} \bar{A}_{32} \\ (-1)^{1+3} \bar{A}_{13} & (-1)^{2+3} \bar{A}_{23} & (-1)^{3+3} \bar{A}_{33} \end{vmatrix} \\
 &= \begin{vmatrix} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} & -\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} & \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} \\ -\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} & \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} & -\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} \\ \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} & -\begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix} & \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \end{vmatrix}
 \end{aligned}$$

38. If A^* is the adjoint of A , then $AA^* = A^*A = |A|$.

PROOF:

By Prob. 35,

$$\begin{aligned}
 AA^* &= (a_{ij})(A_{ij})' = \left(\sum_{k=1}^n a_{ik} A_{jk} \right) = (|A| \delta_{ij}) = |A| (\delta_{ij}) \\
 &= |A| \begin{vmatrix} \delta_{11} & \delta_{12} & \dots & \delta_{1n} \\ \delta_{21} & \delta_{22} & \dots & \delta_{2n} \\ \dots & \dots & \dots & \dots \\ \delta_{n1} & \delta_{n2} & \dots & \delta_{nn} \end{vmatrix} = |a| \begin{vmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{vmatrix} = |A| |1| = |A|
 \end{aligned}$$

Likewise, $A^*A = |A|$, i.e. $A^*A = AA^* = |A|$.

39. Prove Th. 4.1.3.2.21.

PROOF:

By Prob. 38, $AA^* = A^*A = |A| |1|$, i.e. $|1| = A(A^*/|A|) = (A^*/|A|)A$ if $|A| \neq 0$. Hence $B = A^*/|A|$ such that $AB = BA = |1|$, i.e. $B = A^{-1}$.

A^{-1} is unique; since if there exists C , $C \neq B$, such that $AC = CA = |1|$, then $C = |1|C = (A^{-1}A)C = A^{-1}(AC) = A^{-1}|1| = A^{-1} = B$.

40. Find $|A|$, A^* , and A^{-1} , given a square matrix: $A = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 3 & -2 \\ -4 & -6 & 3 \end{bmatrix}$.

Solution:

$$|A| = 26, \quad \text{and} \quad A^* = \begin{bmatrix} \begin{vmatrix} 3 & -2 \\ -6 & 3 \end{vmatrix} & -\begin{vmatrix} 1 & 1 \\ -6 & 3 \end{vmatrix} & \begin{vmatrix} 1 & 1 \\ 3 & -2 \end{vmatrix} \\ -\begin{vmatrix} -1 & -2 \\ -4 & 3 \end{vmatrix} & \begin{vmatrix} 1 & 1 \\ -4 & 3 \end{vmatrix} & -\begin{vmatrix} 1 & 1 \\ -1 & -2 \end{vmatrix} \\ \begin{vmatrix} -1 & 3 \\ -4 & -6 \end{vmatrix} & -\begin{vmatrix} 1 & 1 \\ -4 & -6 \end{vmatrix} & \begin{vmatrix} 1 & 1 \\ -1 & 3 \end{vmatrix} \end{bmatrix} = \begin{bmatrix} -3 & -9 & -5 \\ 11 & 7 & 1 \\ 18 & 2 & 4 \end{bmatrix}$$

$$\text{Hence } A^{-1} = A^*/|A| = 1/26 \cdot \begin{bmatrix} -3 & -9 & -5 \\ 11 & 7 & 1 \\ 18 & 2 & 4 \end{bmatrix}.$$

41. If A and B are nonsingular, then

$$(i) (AB)^* = B^*A^*, \quad (ii) (AB)^{-1} = B^{-1}A^{-1}$$

PROOF:

(i) By Prob. 38, $AB(AB)^* = |AB| |1|$, which implies, by Prob. 33,

$$A^*AB(AB)^* = A^*|AB| |1|$$

which in turn implies

$$|A|B(AB)^* = A^*|A| |B|$$

from which it follows, dividing both sides by $|A| \neq 0$,

$$B(AB)^* = A^*|B|$$

Repeating the process, $B(AB)^* = A^*|B| \rightarrow B^*B(AB)^* = B^*A^*|B| \rightarrow |B| |1|(AB)^* = B^*A^*|B| \rightarrow (AB)^* = B^*A^*$.

(ii) Likewise, $(AB)^{-1}AB = |1| \rightarrow (AB)^{-1}ABB^{-1} = |1|B^{-1} \rightarrow (AB)^{-1}A|1| = B^{-1} \rightarrow (AB)^{-1}AA^{-1} = B^{-1}A^{-1} \rightarrow (AB)^{-1}|1| = B^{-1}A^{-1} \rightarrow (AB)^{-1} = B^{-1}A^{-1}$. (Or what is the same, $AB(AB)^{-1} = |1| \rightarrow (AB)^{-1} = B^{-1}A^{-1}$.)

Note. Both results can be readily generalized to

$$(i) (A_1 A_2 \dots A_n)^* = A_n^* A_{n-1}^* \dots A_1^*, \quad (ii) (A_1 A_2 \dots A_n)^{-1} = A_n^{-1} A_{n-1}^{-1} \dots A_1^{-1}$$

if A_1, A_2, \dots, A_n are nonsingular.

42. Prove Th. 4.1.3.2.22.

PROOF:

\bar{A} satisfies **R1-8** as follows: **R1-5**, by Th. 4.1.3.2.11; **R6**, by Df. 4.1.3.2.12; **R7**, by Prob. 17; **R8**, by Prob. 18. Since multiplication is not commutative in \bar{A} (cf. Df. 4.1.3.2.12), \bar{A} is thus a noncommutative ring.

Chapter 4.2

*Subrings

*§4.2.1 Subrings in General

Df. 4.2.1.1 A ring X is said to be *embedded* in a ring Y if Y contains a subring X' isomorphic to X . Y is then called an *extension* (cf. Df. 5.3.1.5) of X .

In general, a set S is embedded in a set R if S is a subset of R while the operative rules for the elements of S are the same whether these elements are considered in S or R .

Example:

As in Th. 4.1.2.3.8, the integral domain I of all integers can be embedded as a subdomain in a field Q , each element of which is a quotient of integers of I ; cf. also Th. 4.1.1.10-11.

Th. 4.2.1.2 Any ring R can be embedded in a ring R' with unity. (Cf. Prob. 1.)

Such an embedding theorem as above is to prove the existence of an algebraic structure with prescribed properties which contains a substructure isomorphic to a given structure, as is exemplified in the following theorems.

Th. 4.2.1.3 The set M of all multiples of the unity 1 of an integral domain D is a minimal subdomain of D . M is then isomorphic to the set I of all integers if D has characteristic zero while it is isomorphic to the set I' of integers modulo p , a prime, if D has characteristic p . (Cf. Prob. 2.)

This theorem may be slightly modified into the following form:

Th. 4.2.1.4 A field F_1 of characteristic zero contains a subfield F'_1 isomorphic to the field R of rational numbers, and a field F_2 of characteristic p , a prime, contains a subfield F'_2 isomorphic to the ring P of integers modulo p . (Cf. Th. 4.1.2.3.8, and Prob. 3 below.)

This theorem implies that every field contains a unique subfield which contains no proper subfields; i.e. every field contains one, and only one *prime field* (cf. Df. 4.1.2.4.2b and Supplementary Prob. 3.25).

The prime field can be defined to be the meet (or intersection) of all subfields of a field F , since the intersection of any number of subfields is again a subfield, which is no doubt the smallest (cf. Prob. 4 below); the only prime fields of F , are then the field R of rational numbers and the field P of integers modulo p .

Th. 4.2.1.5 For every integral domain D there exists a field Q containing a subdomain D' isomorphic to D (cf. Th. 4.1.2.3.8), which is the set of all quotients of the form: a'/b' , where $a', b' \in D'$ and $b' \neq 0$. (Cf. Prob. 5.)

Th. 4.2.1.6 If F is a field which contains a subdomain D'' isomorphic to an integral domain D , then the set F' of all quotients of the form: a''/b'' , where $a'', b'' \in D''$, $b'' \neq 0$, is a subfield isomorphic to Q , obtained by Th. 4.2.1.5. (Cf. Prob. 6.)

The results, obtained by Th.4.2.1.5-6 (and Th.4.1.1.11), make it now self-evident that every integral domain D can be embedded in a field, and actually is contained in a quotient field Q which is the smallest field containing D . If D is here equated to be the domain I of integers, then the field R of rational numbers can be actually constructed from the integers simply by defining R to be the quotient field of I (cf. §5.1.1, Prob. 1-4).

Solved Problems

1. Prove Th.4.2.1.2.

PROOF:

Let I be the ring of integers and R' be the cartesian product (cf. Df.2.2.2.3) of I by R , i.e. $R' = I \times R = \{(a, b)\}$, where $a \in I$, $b \in R$, and $(a_1, b_1) = (a_2, b_2)$ iff $a_1 = a_2$ and $b_1 = b_2$, where $a_i \in I$ and $b_i \in R$, $i = 1, 2, \dots$

Define, then, addition and multiplication in R' :

$$\mathbf{R}'1: (a_1, b_1) + (a_2, b_2) = (a_1 + a_2, b_1 + b_2)$$

$$\mathbf{R}'6: (a_1, b_1)(a_2, b_2) = (a_1a_2, a_1b_2 + a_2b_1 + b_1b_2)$$

Other properties in R' are then found as follows:

$$\begin{aligned} \mathbf{R}'2: (a_1, b_1) + ((a_2, b_2) + (a_3, b_3)) &= (a_1 + (a_2 + a_3), b_1 + (b_2 + b_3)) = ((a_1 + a_2) + a_3, (b_1 + b_2) + b_3) \\ &= ((a_1, b_1) + (a_2, b_2)) + (a_3, b_3) \end{aligned}$$

$$\mathbf{R}'3: (0, 0)$$

$$\mathbf{R}'4: -(a_1, b_1) = (-a_1, -b_1)$$

$$\mathbf{R}'5: (a_1, b_1) + (a_2, b_2) = (a_1 + a_2, b_1 + b_2) = (a_2 + a_1, b_2 + b_1) = (a_2, b_2) + (a_1, b_1)$$

$$\begin{aligned} \mathbf{R}'7: (a_1, b_1)((a_2, b_2)(a_3, b_3)) &= (a_1, b_1)(a_2a_3, a_2b_3 + a_3b_2 + b_2b_3) \\ &= (a_1(a_2a_3), a_1((a_2b_3 + a_3b_2 + b_2b_3) + b_1(a_2a_3) + b_1(a_2b_3 + a_3b_2 + b_2b_3)) \\ &= ((a_1a_2)a_3, (a_1a_2)b_3 + (a_1b_2 + a_2b_1 + b_1b_2)a_3 + (a_1b_2 + a_2b_1 + b_1b_2)b_3) \\ &= (a_1a_2, a_1b_2 + a_2b_1 + b_1b_2)(a_3, b_3) = ((a_1, b_1)(a_2, b_2))(a_3, b_3) \end{aligned}$$

$$\begin{aligned} \mathbf{R}'8: (a_1, b_1)((a_2, b_2) + (a_3, b_3)) &= (a_1, b_1)(a_2 + a_3, b_2 + b_3) \\ &= (a_1(a_2 + a_3), a_1(b_2 + b_3) + b_1(a_2 + a_3) + b_1(b_2 + b_3)) \\ &= (a_1a_2 + a_1a_3, (a_1b_2 + a_1b_3) + (a_2b_1 + a_3b_1) + (b_1b_2 + b_1b_3)) \\ &= (a_1a_2, a_1b_2 + a_2b_1 + b_1b_2) + (a_1a_3, a_1b_3 + a_3b_1 + b_1b_3) \\ &= (a_1, b_1)(a_2, b_2) + (a_1, b_1)(a_3, b_3) \end{aligned}$$

Likewise, $((a_1, b_1) + (a_2, b_2))(a_3, b_3) = (a_1, b_1)(a_3, b_3) + (a_2, b_2)(a_3, b_3)$.

$$\mathbf{R}'10: (1, 0)$$

Hence R' is a ring with unity.

Now, let a subset of R' be R'_1 , whose elements are of the form $(0, b)$; then, since

$$(0, b_1) + (0, b_2) = (0, b_1 + b_2), \quad (0, b_1)(0, b_2) = (0, b_1b_2), \quad 0 = (0, 0), \quad -(0, b_1) = (0, -b_1)$$

R'_1 is a subring of R' .

Furthermore, the substitution $b'_i = (0, b_i)$ sets up a 1-1 mapping: $b_i \leftrightarrow b'_i$, which reveals an isomorphism of R into R'_1 . Hence R is now embedded in R' .

Note. As is evident in the above context, the set I of integers is also embedded in R' , since the mapping: $a_i \leftrightarrow (a_i, 0)$ is an isomorphism of I into a subring R'_2 of R' .

2. Prove Th. 4.2.1.3.

PROOF:

- (i) Let M be the set of all multiples of the form $n \cdot 1$, $n \in I$, of the multiplicative identity 1 of the integral domain D . Then it immediately follows that $a \cdot 1, b \cdot 1 \in M$ implies

$$\begin{aligned} a \cdot 1 + b \cdot 1 &= (a+b) \cdot 1 \in M, & (a \cdot 1)(b \cdot 1) &= (ab) \cdot 1 \in M, \\ 0 \cdot 1 &= 1 \in M, & -(a \cdot 1) &= -a \cdot 1 \in M \end{aligned}$$

Hence M is a subdomain of D .

Furthermore, since any subdomain D' of D must contain 1 of D , and since D' is closed under addition, D' must contain all multiples of 1. Hence $M \subseteq D'$, proving that M is the least subdomain of D .

- (ii) Let H be the homomorphic mapping of I onto M , i.e.

$$H: I \rightarrow M, \quad n \mapsto n \cdot 1$$

Then, if D has characteristic zero, H is a 1-1 mapping of I into M , since $a \cdot 1 = b \cdot 1$, $a \neq b$, implies $(a-b) \cdot 1 = 0$, which in turn implies $a-b=0$, i.e. $a=b$, since the characteristic of 1 itself is zero. Hence M is isomorphic to I .

- (iii) If D is of a prime characteristic p , then it is first to be proved that, for every $a \neq 0$, $a \cdot 1 = 0$ implies $p \mid a$.

Since p is a prime, $(p, a) = 1$ or p , i.e. by Df. 4.1.2.2.18,

$$(p, a) = rp + sa, \quad r, s \in I$$

But, if $a \cdot 1 = 0$, then $(p, a) \cdot 1 = (rp) \cdot 1 + (sa) \cdot 1 = r(p \cdot 1) + s(a \cdot 1) = 0$, which implies $(p, a) = p$, since $1 \neq 0$. Hence $p \mid a$.

It is to be proved, secondly, that the mapping F , viz.

$$F: I \rightarrow M, \quad n \mapsto n \cdot 1$$

where $n \in I$ and $\{n\}$ is a subset of the residue class modulo p , is 1-1, i.e. that $a \cdot 1 = b \cdot 1$ iff $\{a\} \equiv \{b\} \pmod{p}$.

If $\{a\} \equiv \{b\} \pmod{p}$, then $a-b = p \cdot q$, $q \in I$, i.e. $(a-b) \cdot 1 = (p \cdot q) \cdot 1 = q \cdot (p \cdot 1) = 0$, which implies $a \cdot 1 = b \cdot 1$.

Conversely, if $a \cdot 1 = b \cdot 1$, then $(a-b) \cdot 1 = 0$, and since, by Df. 4.1.2.2.19, $a-b = p \cdot q$ for some $q \in I$, it follows, by Df. 4.1.2.2.20, that $a \cdot 1 = b \cdot 1$ implies $\{a\} \equiv \{b\} \pmod{p}$.

Furthermore, since $\{a\} + \{b\} \leftrightarrow (a \cdot 1) + (b \cdot 1)$ and $\{a\} \cdot \{b\} \leftrightarrow (a \cdot 1) \cdot (b \cdot 1)$, the mapping F is an isomorphism of M into the residue class I' modulo p , which verifies that I' is a subdomain of D of characteristic p .

This completes the proof.

3. Prove Th. 4.2.1.4.

PROOF:

Since F_1 is *a fortiori* an integral domain, the set I of integers is the least subfield of F_1 , by Th. 4.2.1.3 above. Also $n^{-1} \in F_1$ for every $n \in I$, $n \neq 0$, since F_1 is a field. Hence F_1 contains a subfield F_1' isomorphic to the quotient field Q (cf. Th. 4.1.2.3.7), which in turn is isomorphic to the set R of rational numbers (cf. Th. 4.1.2.3.9). I being the least subdomain of F_1 , R is then evidently the least subfield of F_1 .

The rest is proved by two isomorphisms: $F_2 \leftrightarrow D$ and $P \leftrightarrow I'$, D and I' being defined by Th. 4.2.1.3 above.

4. The meet of any number of subfields of a field F is again a subfield of F .**PROOF:**

Let $M = \cap S_i$, $i = 1, 2, \dots, n$, where any of S_i is a subfield of F ; then M contains, by definition, all elements which are contained in any of S_i , where all of F1-11 are satisfied, again by definition. Hence M contains 0 and 1, and if $a, b \in M$, then $a+b$, $-a$, $a \cdot b$, $a^{-1} \in M$, satisfying Df. 4.1.2.4.2a.

M is thus a subfield of F .

5. Prove Th. 4.2.1.4.

PROOF:

Since Q has already been proved, by Th. 4.1.2.3.7, to be a field, let D' be a subset of Q whose elements are of the form (a, e) , where $a \in I$ and e is the unity of D . Then the mapping M :

$$M(a) \leftrightarrow (a, e)$$

is an isomorphism (cf. Th. 4.2.1.3, ii) of D' into D .

Hence D' is an integral domain.

Furthermore, $D' \subseteq Q$ implies $a'/b' \in Q$, where $a', b' \in D'$ and $b' \neq 0$, and moreover, since any element $(a, b) \in Q$ is the solution $(a, e)/(b, e)$ of an equation $(b, e)(x, y) = (a, e)$ in D' , every element of Q is a quotient of the form a'/b' .

6. Prove Th. 4.2.1.6.

PROOF:

Since, by hypothesis, D is isomorphic to $D'' \subset F$, let the isomorphism be defined by the mapping M , for every $a \in D$:

$$M(a) \leftrightarrow a'' \in D''$$

Then the mapping M' , defined by

$$M'((a, b)) \leftrightarrow M(a)/M(b) \leftrightarrow a''/b'' \in F'$$

for all $(a, b) \in Q$, is also an isomorphism, viz. of F' into Q , where F' is obviously a subfield of F .

*§4.2.2 Ideals

Df. 4.2.2.1 An ideal \bar{I} in a ring R is an additive subgroup of R with the closure property that $a \in \bar{I}$ and $r \in R$ imply $ar, ra \in \bar{I}$.

Since $ar = ra$ in a commutative ring, the first definition may be modified as follows:

Df. 4.2.2.1a An ideal \bar{I} in a commutative ring R is an additive subgroup of R with the closure property that $a \in \bar{I}$ and $r \in R$ imply $ar \in \bar{I}$ (or what is the same, $ra \in \bar{I}$).

Example:

The set E of even integers in the ring I of integers is an ideal (cf. Prob. 3); so is the set M of all multiples of an integer, say 5, in I , or indeed the ring I in I itself, as can be verified without difficulty.

It must be emphasized, however, that these two definitions yield no ideals other than those in a (commutative) ring; ideals in an algebraic number field, for instance, are defined otherwise.

Ideals in a ring in general may be defined also in terms of cosets (cf. Df. 3.2.2.2), viz.

Df. 4.2.2.1b An ideal \bar{I} is a *subring* of a ring R if $r\bar{I} \subseteq \bar{I}$ and $\bar{I}r \subseteq \bar{I}$ for every $r \in R$.

If ideals are defined by this definition, then the following definition of ideals becomes deducible, viz.

Th. 4.2.2.2 A complex C of a ring R is an ideal iff $(a-b), ar, ra \in C$ for every $a, b \in C$ and $r \in R$. (Cf. Prob. 2.)

Df. 4.2.2.1b and Th. 4.2.2.2 can be simplified, like Th. 4.2.2.1a, if R is defined to be commutative. Since a ring is not always commutative under multiplication, however, it is possible, though seldom practiced in the following pages, to consider one-sided ideals as follows:

Df. 4.2.2.3 An ideal \bar{I} is a *left-ideal* if $a \in \bar{I}$ and $r \in R$, as in Th. 4.2.2.1, imply only $ra \in \bar{I}$, and a *right-ideal* if they imply only $ar \in \bar{I}$. Ideals in a commutative ring are then called, in the same context, *two-sided ideals* or, as below, simply *ideals*.

The concept of one-sided ideals is not superfluous, since it may be employed, for instance, to characterize sfields (cf. Supplementary Prob. 4.58).

As has already been observed, the whole ring R , like I in the first example, is always an ideal, just as any group is a subgroup of itself. Of rings as such, the whole ring R and the minimal ring 0 are specially defined as follows:

Df. 4.2.2.4 The whole set R and the set 0 of the zero element alone of a ring R are called the *improper* (or *trivial*) *ideals*, while all other ideals are called *proper*. Of the two trivial ideals R and 0 , the latter is sometimes called the *zero ideal* and any other ideal a *nonzero ideal*.

These trivial ideals may not be literally trivial, since they may be able to characterize some important sets, e.g.

Th. 4.2.2.5 A sfield F^* has only the two trivial ideals; so does a field F . (Cf. Prob. 7.)

More significantly, the ring I of integers also has a special name for its special rôle in the theory of ideals, viz.

Df. 4.2.2.6 The ring I of integers is called the *unit ideal*, denoted by (1) , if it is to be considered an ideal in I itself or generally in a commutative ring R with unity; i.e. $I = (1)$. (Cf. Prob. 8-9.)

I is obviously an ideal generated by the unity itself, and every element of I is thus a multiple of 1. More generally, the set M of all multiples of an integer n , denoted by (n) , is also an ideal in I (cf. Prob. 4-5); hence the following definition.

Df. 4.2.3.7 An ideal, each element of which is a multiple of an element a of a commutative ring R with unity, is said to be *generated* by a , called a *principal ideal*, and denoted by (a) .

I is thus a principal ideal, which is further characterized as follows:

Th. 4.2.2.8 Every ideal in I is principal. (Cf. Prob. 5.)

This theorem is actually a restatement of the Division Algorithm for integers, and the similar algorithm for polynomials over a field F (cf. Th. 5.3.1.2) will yield a similar theorem:

Th. 4.2.2.9 Every ideal in the domain $F(x)$ of polynomials over a field F is principal. (Cf. Prob. 6.)

Since ideals in a ring are not always principal, the rings of Th. 4.2.2.8-9 are in a class by themselves, viz.

Df. 4.2.2.10 A commutative ring R is called a *principal ideal ring* if R has the property that every ideal in R is a principal ideal.

Besides I and $F(x)$, given above, there are of course other principal ideal rings; e.g. the ring obtained by the following definition:

Df. 4.2.2.11 A ring E is called *Euclidean* if a nonnegative integer $\bar{n}(a)$ can be assigned to $a \in E$ such that (i) $\bar{n}(ab) \geq \bar{n}(a)$ for $b \in E$ and $ab \neq 0$, and (ii) there always exist $q, r \in E$ such that $b = qa + r$, for any $a \neq 0, b \in E$, where either $\bar{n}(a) > \bar{n}(r)$ or $r = 0$.

E is a principal ideal ring (cf. Prob. 9), which is in fact an analog of the Euclidean Algorithm (cf. §4.1.2.3, Prob. 31), generalized to arbitrary rings.

Since the principal ideal (a) generated by an element a of a ring R generally contains all elements of the form $ra + na$, where $r, a \in R$ and $n \in I$ (cf. Prob. 8 below), a natural extension is the ideal (a_1, a_2, \dots, a_n) in R generated by the finite number of elements $a_1, a_2, \dots, a_n \in R$, which is the set of elements of the form $\sum_i a_i r_i + \sum_j a_j n_j$, $i, j = 1, 2, \dots, n$, where $r_i \in R$ and $n_j \in I$. The elements a_1, a_2, \dots, a_n are then said to form a *basis* of the ideal. (The principal ideal (a) is thus, in this context, an ideal with a basis consisting of only one element a .)

This is further generalized by the sums and products of ideals, defined as follows.

Df. 4.2.2.12 The (*direct*) *sum* of any two ideals A and B is the set $\{a_i + b_j\}$, where $a_i \in A$ and $b_j \in B$, and the *product* of A and B is the set $\{\sum a_i b_j\}$, $i = 1, 2, \dots, m$; $j = 1, 2, \dots, n$.

In general, then, two ideals A and B in a commutative ring R generated by bases $A = (a_1, a_2, \dots, a_m)$ and $B = (b_1, b_2, \dots, b_n)$, imply that their sum is of the form $\sum_i a_i x_i + \sum_j b_j y_j$, where the basis is

$$(a_1, \dots, a_m) + (b_1, \dots, b_n) = (a_1, \dots, a_m, b_1, \dots, b_n)$$

and their product is of the form $(\sum_i a_i x_i)(\sum_j b_j y_j)$, where the basis is

$$(a_1, \dots, a_m) \cdot (b_1, \dots, b_n) = (a_1 b_1, a_1 b_2, \dots, a_m b_{n-1}, a_m b_n)$$

Solved Problems

1. An ideal \bar{I} in a ring R is necessarily a subring of R .

PROOF:

Since \bar{I} is an additive subgroup of R by Df. 4.2.2.1, $a \in \bar{I}$ implies $a - a = 0 \in \bar{I}$, $0 - a = -a \in \bar{I}$, and $a + b = a - (-b) \in \bar{I}$.

Furthermore, by the given closure property, $a, b \in \bar{I}$ implies $ab, ba \in \bar{I}$.

Hence, by Df. 4.1.1.7, \bar{I} is a subring of R .

2. Prove Th. 4.2.2.2.

PROOF:

- (i) If $(a - b) \in C$ and $ar, ra \in C$ for every $a, b \in C$ and $r \in R$, then it immediately follows, from Th. 4.1.1.7, that C is a subring of R .

Hence $rC \subseteq C$ and $Cr \subseteq C$ for every $r \in R$, which proves, by the second definition of Df. 4.2.2.1, that C is an ideal.

- (ii) Conversely, if C is an ideal in a ring R , then, by Prob. 1 above, it is necessarily a subring of R , which implies $(a - b), ar, ra \in C$ for every $a, b \in C$ and $r \in R$, completing the proof.

3. The set E of even integers, which is evidently a group under addition, forms an ideal in the ring I of all integers.

PROOF:

Let $a, b \in E$, where $a = 2a'$ and $b = 2b'$ for every $a', b' \in I$; then $a - b = 2a' - 2b' = 2(a' - b') \in E$. Also, for every $r \in I$, $ar = ra = 2ra' \in E$. Hence, by Df. 4.2.2.2, E is an ideal.

4. The set M of all multiples of an integer n , denoted by (n) , is an ideal in I .

PROOF:

For every $a, b \in M$, where $a = na'$ and $b = nb'$ for any $a', b' \in I$, $a - b = na' - nb' = n(a' - b') \in M$, and also for every $m \in I$, $am = ma = n(a'm) \in M$. Furthermore, (n) is an additive subgroup of I , since $0 = n \cdot 0 \in (n)$ and $-a = -na = n(-a) \in (n)$. Hence (n) is an ideal in I .

5. The ideal (n) , of Prob. 4, is the only ideal in I .

PROOF:

- (i) Assume that a non-empty set L , $L \subseteq I$, is an ideal in I ; then L must have at least one element, say 0, which, however, at once implies that $L = (0)$. If L has only one element other than 0, say n , then manifestly $L = (n)$.
- (ii) Assume that L has more than one nonzero element, then L contains some positive integers ($\because L \subseteq I$, by hypothesis; i.e. L is a ring and contains some additive inverses), where n may denote the least positive integer in L . If m is any other integer in L , then, by Th. 4.1.2.3, there exist $q, r \in L$ such that

$$m = n \cdot q + r, \quad 0 \leq r < n$$

Since L is an ideal and $m \in L$, i.e. $r \in L$ and $n \cdot q \in L$, it follows that $m - n \cdot q = r \in L$ and that $r = 0$ and $m = n \cdot q$, n being the least positive integer such that $r < n$. Hence, again, $L = (n)$.

(i) and (ii) exhausting the cases, the proof is complete.

It must be noted that the theorem above, stated in terms of "principal ideals", yields Th. 4.2.2.8. Note, also, that all ideals in I are thus unit ideals.

6. Prove Th. 4.2.2.9.

PROOF:

- (i) Let F' be an ideal in $F[x]$, and let $F' = 0$, viz. the ideal consisting of 0 alone; then, as in Prob. 5, (i) above, $F' = (0)$, viz. the principal ideal generated by 0, i.e. zero polynomial over F .
- (ii) Let $F' \neq 0$ and $g(x)$ be a non-zero polynomial of the least degree in F' ; then, for every $f(x) \in F'[x]$, there exist $q(x), r(x) \in F[x]$ such that, by Th. 5.3.1.2,

$$f(x) = g(x)q(x) + r(x), \quad 0 \leq \deg r(x) < \deg g(x)$$

i.e. $r(x) = f(x) - g(x)q(x)$, which implies $r(x) = 0$ since $\deg r(x) < \deg g(x)$. Hence $f(x) = g(x)q(x)$, i.e. $f(x) \in (g(x))$, proving that every $f(x) \in F'[x]$ is contained in the principal ideal $(g(x))$. Hence $F' \subseteq (g(x))$.

Furthermore, $g(x) \in F'$, i.e. $(g(x)) \subseteq F'$. Hence $F' = (g(x))$, proving that every ideal in $F[x]$ is principal.

7. Prove Th. 4.2.2.5.

PROOF:

- (i) If A is an ideal in a sfield F^* , then either $A = (0)$ or $A \neq (0)$. If $A \neq (0)$, then let $a \in A$ and $a \neq 0$, which implies, by Df. 4.2.2.1, $a^{-1}a = 1 \in A$, which in turn implies, by the same definition, $b \cdot 1 = b \in A$ for every $b \in F^*$. Hence $A = F^*$, completing the proof.
- (ii) The proof of (i) can be repeated, *a fortiori*, for a field F .

8. If R is a commutative ring with unity and I the set of all integers, then the set S of all elements of the form $ra + na$, where $a, r \in R$ and $n \in I$, is a principal ideal (a) .

PROOF:

Let $s_1, s_2 \in S$, where $s_1 = r_1a + n_1a$, $s_2 = r_2a + n_2a$, and $r_1, r_2 \in R$, $n_1, n_2 \in I$. Then S is an ideal in R , since S is evidently an additive subgroup of R and furthermore $s_1 - s_2 = (r_1 - r_2)a + (n_1 - n_2)a = r_3a + n_3a = s_3 \in S$ and, for any $R \in R$,

$$rs_1 = r(r_1a + n_1a) = (rr_1 + rn_1)a = r_4a + 0 \cdot a = s_4 \in S$$

S is also a principal ideal (a) , since R has a unity e and

$$ra + na = ra + n(ea) = (r + ne)a = r'a, \quad r \in R$$

9. Every Euclidean ring is a principal ideal ring.

PROOF:

Let E be a Euclidean ring and \bar{I} an ideal in E .

If \bar{I} is the zero ideal (cf. Df. 4.2.2.4), it is trivially principal, and if it is not, then there must exist nonzero elements in \bar{I} , one of which, say a , may be chosen such that $\bar{n}(a) > \bar{n}(c)$ for $c \notin \bar{I}$. Then, by the following reason, \bar{I} is exactly the principal ideal generated by a .

Since E is given as a Euclidean ring, $b = qa + r$ for every $b \in \bar{I}$ and some $q, r \in E$ imply that either $\bar{n}(a) > \bar{n}(b)$ or $r = 0$. The latter follows at once, however, since $\bar{n}(a)$ has already been made minimal. Hence $b = qa$, implying that every element b in \bar{I} is thus a multiple of the generating element a and that $\bar{I} \subseteq (a)$, while $(a) \subseteq \bar{I}$, since $a \in \bar{I}$. Hence $\bar{I} = (a)$, completing the proof.

*§4.2.3 Quotient Rings

Df. 4.2.3.1 If R is a ring and M an ideal in R , then the set Q of elements of the form $r + m$, for a fixed element $r \in R$ and every $m \in M$, is called a *residue class* (or *remainder class*) in R .

The set Q , viz. $r + M$, is evidently a *coset* in terms of an additive group, just as gG_i , $g \in G$, is a coset in group theory (cf. Df. 4.2.2.2). In the same sense, two residue classes $r_1 + M$ and $r_2 + M$, where $r_1, r_2 \in R$, may be interpreted in terms of modules (i.e. additive Abelian groups) and considered equivalent, viz. $r_1 \equiv r_2 \pmod{M}$ iff $r_1 - r_2 \in M$. (Cf. Prob. 1.) This is a generalization of congruences, formulated in terms of ideals as follows.

Th. 4.2.3.2 If R is a commutative ring, M an ideal in R , and every $a, b, c, d \in R$, then $a \equiv b \pmod{M}$ and $c \equiv d \pmod{M}$ imply (i) $a + c \equiv b + d \pmod{M}$, (ii) $ac \equiv bd \pmod{M}$, and (iii) $ar \equiv br \pmod{M}$ for any $r \in R$. (Cf. Prob. 2.)

This theorem indicates that residue classes function in rings in analogy to groups (Th. 3.2.6.8-9). The analogy becomes closer still when the cosets, i.e. residue classes, in a commutative ring R , which partition R into non-overlapping complexes of R (cf. Th. 3.2.2.6), actually form a ring.

Th. 4.2.3.3 The set Q of residue classes of an ideal A in a commutative ring R is a ring. (Cf. Prob. 3.)

It is but natural at this juncture to call the set Q of Df. 4.2.3.1 a *quotient ring*, analogous to a quotient group, since an ideal A in a commutative ring R is a normal subgroup of the additive group of R , which implies, as has already been seen, that the elements of R can be separated into cosets of A in R , viz. residue classes of R modulo A , an ideal. Hence the following restatement of Df. 4.2.3.1.

Df. 4.2.3.1a The residue class Q of elements of the form $r + a$, for a fixed element $r \in R$, a ring, and every $a \in A$, an ideal in R , is called a *quotient ring* (or *factor ring*), denoted by R/A .

This notation is plainly in accordance with the quotient group G/G_i of G by G_i (or what is the same, the factor group of G_i in G) in group theory (cf. Df. 3.2.5.2), where G_i is a normal subgroup of G .

The analogy goes further, as is obvious in the following theorems of homomorphisms.

Th. 4.2.3.4 The set S of elements mapped onto zero in any homomorphism of a ring R is an ideal in R . (Cf. Prob. 6.)

This theorem, as well as others below, articulates a functional similarity between normal subgroups in a group (cf. Th. 3.2.5.8) and ideals in a ring, even up to the “kernel” (cf. Df. 3.2.5.9) of homomorphism, which is S in the above context, just as C in the following theorem:

Th. 4.2.3.5 If a homomorphism H of a ring R onto a ring R' is a mapping of a complex C of R onto the zero element of R' , then C is an ideal in R and R/C is isomorphic to R' . (Cf. Prob. 7, and also cf. Th. 3.2.6.18.)

Sometimes, parallel to proper and improper ideals, homomorphic images may be classified as follows:

Df. 4.2.3.6 A homomorphism which maps a proper ideal in a ring or the whole ring onto zero is called *proper*; otherwise, it is called *improper*.

It follows at once, then, that a sfield F^* or a field F has no proper homomorphic images, since F^* or F has no proper ideals (cf. Th. 4.2.2.5).

* * * * *

In relation to quotient rings, two special types of ideals may be studied separately.

Df. 4.2.3.7 An ideal P in a ring R is called a *prime ideal* if $ab \in P$ implies either $a \in P$ or $b \in P$.

Example:

The principal ideal (3) is prime, since $ab \in (3)$ implies that a or b must be a multiple of 3, while (10) is not prime, since e.g. $20 \in (10)$, yet $20 = 4 \cdot 5$ implies $4 \notin (10)$ and $5 \notin (10)$.

Th. 4.2.3.8 The quotient ring $Q = R/P$, where P is an ideal in a commutative ring R with unity, is an integral domain iff P is prime. (Cf. Prob. 8.)

Example:

The quotient ring $R/(10)$ cannot form an integral domain, since some nonzero elements such as $(2 + (10))$ and $(5 + (10))$ may turn out to be zero-divisors (since $(2 + (10)) \cdot (5 + (10)) = 10 + (10) = (10)$).

As a matter of fact, every element which is neither zero nor prime (cf. Df. 4.1.2.3.12) in an integral domain, where the unique factorization theorem (cf. Th. 4.1.2.3.17) holds, generates a nonprime ideal (cf. Supplementary Prob. 4.57).

Example:

The ideals $(3), (5), (7)$, etc. are all prime in I ; so is $(x^2 + 2)$ in $\bar{R}\{x\}$, of polynomials over the field \bar{R} of real numbers, although it is not a prime ideal in $C\{x\}$, of polynomials over the field C of complex numbers, since $x^2 + 2 = (x + \sqrt{2}i)(x - \sqrt{2}i)$ is certainly not a prime element in $C\{x\}$.

The concept of prime ideals may be further articulated through maximal ideals, defined parallel to maximal normal subgroups (cf. Df. 3.2.7.1).

Df. 4.2.3.9 An ideal M in a ring R is called *maximal* if an ideal M' which properly contains M cannot be properly contained in R .

Example:

The principal ideal (4) in the ring I of integers is not maximal, since (4) is properly contained in (2) , which in turn is properly contained in $(1) = I$; the principal ideal (3) , however, is maximal, since it is properly contained only in $(1) = I$.

Th. 4.2.3.10 The quotient ring $Q = R/M$, where R is a commutative ring with unity and M is an ideal in R , is a field iff M is a maximal ideal in R . (Cf. Prob. 10.)

Solved Problems

1. Given an ideal M in a commutative ring R and $r_1, r_2 \in R$, prove that $r_1 \equiv r_2 \pmod{M}$ iff $r_1 - r_2$ is in M .

PROOF:

Since any ideal, say M , in a commutative ring R satisfies the definition of a normal subgroup (cf. Df. 3.2.5.3) under addition (i.e. instead of “ \circ ” or “ \ast ” multiplication under which Df. 4.2.5.3 is defined), and since, by Df. 4.2.3.1, a residue class modulo an ideal M can partition R into disjoint cosets of M in R , the cosets are now given under addition, i.e. of the form $r_i + M$, where r_i is any element of R .

Furthermore, since $aG_i = bG_i$ in a subgroup G_i of a group G (under multiplication), to which a and b belong, iff $a^{-1}b \in G_i$, or what is virtually the same in this context, iff $ab^{-1} \in G_i$, it can be translated in terms of addition, viz. $a + G_i = b + G_i$ iff $a - b \in G_i$. Translate, then, G_i in G into M , an ideal, in R and $a, b \in G$ into $r_1, r_2 \in R$, and the proof is complete.

Note. For example, $M = (2)$ in a ring C of complex numbers of the form $a + bi$, where $a, b \in I$ and $i = \sqrt{-1}$, implies the following four distinct residue classes, viz. $(2), 1 + (2), i + (2), 1 + i + (2)$, all of which may be represented by their general form $a + bi \equiv c + di \pmod{(2)}$, where $0 \leq c, d < 2$.

2. Prove Th. 4.2.3.2.

PROOF:

(This is merely a restatement of Th. 3.2.6.8-9 in terms of modulo ideals (instead of integers).)

- (i) Since, by hypothesis, $(a - b), (c - d) \in M$, it immediately follows that $(a - b) + (c - d) = (a + c) - (b + d) \in M$, i.e. $a + c \equiv b + d \pmod{M}$.
- (ii) Since, as above, $(a - b), (c - d) \in M$, it follows that $c(a - b), (c - d)b \in M$ and $c(a - b) + (c - d)b = ac - bd \in M$, i.e. $ac \equiv bd \pmod{M}$.
- (iii) $(a - b) \in M$ implies $(a - b)r \in M$ for every $r \in R$, which in turn implies $ar - br \in M$, i.e. $ar \equiv br \pmod{M}$.

3. Prove Th. 4.2.3.3.

PROOF:

The two binary operations in the set Q of residue classes are defined, for an ideal A in a commutative ring R and $a, b, c \in R$,

$$\mathbf{Q1:} \quad (a + A) + (b + A) = (a + b) + A \pmod{A},$$

$$\mathbf{Q6:} \quad (a + A) \cdot (b + A) = (a \cdot b) + A \pmod{A},$$

both of which are well-defined, since $a + A = c + A$ and $b + A = d + A$ imply $ab + A = cd + A$ (i.e. $ab - cd \in A$, since $a = c + x$ and $b = d + x'$, $x, x' \in A$, imply $ab = (c + x)(d + x') = cd + (cx' + dx + xx') = cd + x'' \in A$).

$$\begin{aligned} \mathbf{Q2:} \quad (a + A) + ((b + A) + (c + A)) &= (a + A) + ((b + c) + A) = (a + (b + c)) + A \\ &= ((a + b) + c) + A = ((a + b) + A) + (c + A) \\ &= ((a + A) + (b + A)) + (c + A) \end{aligned}$$

$$\mathbf{Q3:} \quad A$$

$$\mathbf{Q4:} \quad -(a + A) = (-a) + A$$

$$\mathbf{Q5:} \quad (a + A) + (b + A) = (a + b) + A = (b + a) + A = (b + A) + (a + A)$$

$$\begin{aligned} \mathbf{Q7:} \quad (a + A)((b + A)(c + A)) &= (a + A)(bc + A) = a(bc) + A = (ab)c + A \\ &= (ab + A)(c + A) = ((a + A)(b + A))(c + A) \end{aligned}$$

$$\begin{aligned} \mathbf{Q8:} \quad (a + A)((b + A) + (c + A)) &= (a + A)((b + c) + A) = a(b + c) + A = (ab + ac) + A \\ &= (ab + A) + (ac + A) = (a + A)(b + A) + (a + A)(c + A) \end{aligned}$$

which completes the proof.

4. There exists an isomorphism between the set I of integers and the residue classes of $I[x]$ modulo M , where $I[x]$ is a polynomial ring over I while M is an ideal $(x-1)$, consisting of all elements of the form $a(x)(x-1)$ for any $a(x) \in I[x]$.

PROOF:

Since $f(x) = (x-1)q(x) + f(1)$ for any element $f(x) \in I[x]$, by Th. 4.2.2.9, it follows at once that $f(x) = f(1) \pmod{(x-1)}$, where obviously $f(1) \in I$, which implies that an integer may determine a residue class.

Furthermore, every integer represents a distinct residue class. For $a \equiv b \pmod{(x-1)}$, i.e. $a - b \in (x-1)$, implies $a - b = (x-1)q(x)$, which in turn implies $\deg(a - b) = \deg((x-1)q(x)) > 0$ unless $q(x) = 0$. Hence it must be the case that $q(x) = 0$, which implies $a - b = 0$, i.e. $a = b$.

Thus two elements of the residue classes of $I[x]$ modulo $(x-1)$, say $a + (x-1)$ and $b + (x-1)$, are distinct, which consequently yield two distinct mappings: $a + (x-1) \leftrightarrow a$ and $b + (x-1) \leftrightarrow b$.

This mapping is isomorphic, since

$$\begin{aligned} (a + (x-1)) + (b + (x-1)) &= (a + b) \pmod{(x-1)} \leftrightarrow a + b, \\ (a + (x-1)) \cdot (b + (x-1)) &= ab \pmod{(x-1)} \leftrightarrow ab \end{aligned}$$

which completes the proof.

Note. The residue classes of $I[x]$ modulo M , the ideal $(x-1)$, is by Df. 4.2.3.3 a quotient ring, of course, and may be denoted by $I[x]/(x-1)$, which is thus isomorphic to I . Cf. Prob. 5 below.

5. The quotient ring $I[x]/(x^2 + 1)$, where $I[x]$ is a polynomial ring over I and $(x^2 + 1)$ denotes an ideal in $I[x]$, is isomorphic to the field C of complex numbers.

PROOF:

Parallel to Prob. 4 above, every polynomial $f(x) \in I[x]$ is now expressed in the form

$$f(x) = (x^2 + 1)q(x) + bx + a, \quad a, b \in I$$

which implies that every integral pair (a, b) , analogous to every integer a in Prob. 4 above, may be proved likewise to determine a distinct residue class.

Thus two elements of the quotient ring $I[x]/(x^2 + 1)$, say $a + bx + q(x)(x^2 + 1) = a'$ and $c + dx + q'(x)(x^2 + 1) = b'$, are distinct, and $a + bx = a' \pmod{(x^2 + 1)}$ and $b' = c + dx \pmod{(x^2 + 1)}$ establish the desired isomorphic mapping of $I[x]/(x^2 + 1)$ into C , since $a + bx \pmod{(x^2 + 1)} \leftrightarrow (a, b)$ and $c + dx \pmod{(x^2 + 1)} \leftrightarrow (c, d)$ imply the isomorphic mappings M and M' :

$$\begin{aligned}
M: \quad (a+bx) + (c+dx) &= (a+c) + (b+d)x \pmod{(x^2+1)} \\
&\Leftrightarrow (a,b) + (c,d) = (a+bi) + (c+di) = (a+c) + (b+d)i = (a+c, b+d) \in C, \\
M': \quad (a+bx) \cdot (c+dx) &= ac + (ad+bc)x + bdx^2 = ac - bd + (ad+bc)x + (bd(x^2+1)) \\
&= (ac-bd) + (ad+bc)x \pmod{(x^2+1)} \\
&\Leftrightarrow (a,b) \cdot (c,d) = (a+bi)(c+di) = (ac-bd) + (ad+bc)i = (ac-bd, ad+bc) \in C,
\end{aligned}$$

completing the proof.

6. Prove Th. 4.2.3.4.

PROOF:

Let H be the given homomorphism; then, by hypothesis,

$$H(s) = o'$$

for every $s \in S$, $S \subseteq R$, where $H(R) = R'$ and $o' \in R'$. Hence, for every $r \in R$, $s_1, s_2 \in S$,

$$\begin{aligned}
H(rs_1) &= H(r)H(s_1) = H(r) \cdot o' = o' \\
H(s_1r) &= H(s_1)H(r) = o' \cdot H(r) = o' \\
\text{and} \quad H(s_1 - s_2) &= H(s_1) - H(s_2) = o' - o' = o'
\end{aligned}$$

since, by hypothesis, $H(s_1) = H(s_2) = o'$.

Hence, by Df. 4.2.3.2, S is an ideal in R .

7. Prove Th. 4.2.3.5.

PROOF:

By Th. 4.2.3.3, R/C is a ring and, by Th. 4.2.3.4, C is an ideal in R .

Now let

$$H: r + C \rightarrow r'$$

for every $r \in R$ and $H(r) = r'$. Then, since C is an ideal, i.e. a normal subgroup of R under addition, the mapping H can be proved to be 1-1, analogous to Th. 3.2.6.18, viz. two distinct images:

$$H(r_1 + C) \leftrightarrow r'_1 \quad \text{and} \quad H(r_2 + C) \leftrightarrow r'_2$$

where $r_1, r_2 \in R$, $r_1 + C, r_2 + C \subset R/C$, and $r'_1, r'_2 \in R'$, imply

$$H((r_1 + C) + (r_2 + C)) = H(r_1 + r_2 + C) \leftrightarrow (r_1 + r_2)' = r'_1 + r'_2 = H(r_1 + C) + H(r_2 + C)$$

$$\text{and} \quad H((r_1 + C) \cdot (r_2 + C)) = H(r_1 r_2 + C) \leftrightarrow (r_1 r_2)' = r'_1 \cdot r'_2 = H(r_1 + C) \cdot H(r_2 + C)$$

8. Prove Th. 4.2.3.8.

PROOF:

(i) If $R = P$, then $R/P = (0)$, which trivially proves the theorem.

(ii) If $R \neq P$, then R/P contains at least two classes, viz. $a + P$ and $b + P$, where $a, b \in R$.

Now, if R/P is an integral domain, then a and b can be chosen in such a way that $ab \in P$, which implies $(a+P)(b+P) = ab+P = P$, meaning the product of the two classes zero, which in turn implies that, since R/P is here an integral domain, one of the classes is zero, i.e. either $a+P = P$ or $b+P = P$, or what is the same, in the notation of congruence: $a \equiv 0 \pmod{P}$ or $b \equiv 0 \pmod{P}$.

Hence either $a \in P$ or $b \in P$ if $ab \in P$, which immediately implies, by Df. 4.2.3.7, that P is a prime ideal in R .

Conversely, if $(a+P)(b+P) = P$, where P is a prime ideal, then $ab+P = P$ since $(a+P)(b+P) = a+b+P$, which implies $ab \equiv 0 \pmod{P}$, which in turn implies $a \equiv 0 \pmod{P}$ or $b \equiv 0 \pmod{P}$, since P is a prime ideal; thus $ab \in P$ implies $a \in P$ or $b \in P$. Hence either $a+P = P$ or $b+P = P$, proving that $a+P$ or $b+P$ must be the zero element of P if the product of $a+P$ and $b+P$ is zero. Hence P must be an integral domain, completing the proof.

9. The mapping $H: R \rightarrow R/A$, where R is a ring and A an ideal in R , is homomorphic.

PROOF:

Let $H: r \rightarrow r + A$, for every $r \in R$; then, for any $a, b \in R$ and $a + A, b + A \in R/A$,

$$a + b \rightarrow H(a + b) = a + b + A = (a + A) + (b + A) = H(a) + H(b)$$

and

$$a \cdot b \rightarrow H(a \cdot b) = ab + A = (a + A)(b + A) = H(a)H(b)$$

Hence $R \rightarrow R/A$ is a homomorphic mapping, completing the theorem.

10. Prove Th. 4.2.3.10.

PROOF:

- (i) If M is a maximal ideal in R , a commutative ring with unity, then $M \subset R$ implies that $R' = R/M$ has at least two classes and, by Prob. 9 above, there exists a homomorphism $H: R \rightarrow R'$, where $a, b, \dots \in R$ and $a', b', \dots \in R'$ such that $a' = a + M$, $b' = b + M$, ..., and of course $o' = o + M$.

Now, to find the property of F11 (cf. Df. 4.1.2.4.1) in R' , let $a' \neq o'$, which implies $a \notin M$ (since $a \in M$ implies $a' = a + M = M = o'$, contrary to the assumption). Furthermore, M being a maximal ideal and $a \notin M$, any ideal generated by M and a must be the whole ring R itself, which implies that $r = ab + m$, for every $r \in R$, where $b \in R$ and $m \in M$. Since r is to denote any element of R , including the unity e of R , it follows that $e = ab + m$, i.e. $e \equiv ab \pmod{M}$ or what is the same: $(e + M) = (a + M)(b + M)$, which is equivalent, by mapping through H , to $e' = a'b'$, establishing F11 in R' .

- (ii) Conversely, if M is not a maximal ideal in R when R' is a field, there must exist an ideal N such that $M \subset N \subset R$. Then, let $m \in M$ (and $m^{-1} \in R'$, R' being a field) such that $m \notin N$, which implies that, R' being a field, there exists an element $a' \in R'$ such that $a'm' = b'$, where b' is an arbitrary element in R' . Hence, by the mapping H , $am \equiv b \pmod{M}$, which evidently holds since $M \subset N$. But then $m \in M$; hence $am \equiv 0 \pmod{N}$, i.e. $b \equiv 0 \pmod{N}$, which implies $b \in N$ for any $b \in R$, which proves that $N = R$, thus implying that M is a maximal ideal if R' is a field.

Supplementary Problems

Part 4

- *4.1. The set of N1-4 (Peano Axioms) of Df. 4.1.2.3.1 is *categorical* (or *complete*); i.e. any set which satisfies N1-4 is isomorphic to the set N of natural numbers.
- *4.2. The set of N1-4 (Peano Axioms) is *independent*; i.e. none of the four axioms among N1-4 is deducible from the other three.
- 4.3. Addition in N (cf. Df. 4.1.2.3.2) is well-defined; i.e., for every $a, b, c \in N$, the binary operation $a + b = c$ is always possible in N and c is here uniquely determined.
- 4.4. Multiplication in N (cf. Df. 4.1.2.3.3) is well-defined; i.e., for every $a, b, c \in N$, the binary operation of $a \cdot b = c$ exists in N , determining c uniquely.
- 4.5. *Subtraction*, i.e. the inverse operation of addition, in N is well-defined; i.e., for every $a, b, c \in N$, $a - b = c$ is feasible, determining c uniquely, iff $a > b$.
- 4.6. If M is a complex of the set N of natural numbers, then M contains a number a such that $a \leq b$ for any $b \in M$.

- 4.7. Let I be the set of all integers and S be a set which contains the set N of natural numbers; then $I = S$ iff every element of S is defined by a difference between two elements of N .
- 4.8. The set I of integers is unique except for the sets isomorphic to I ; i.e. every minimal ring containing the set N of natural numbers is isomorphic to I .
- 4.9. If a ring R contains the set N of natural numbers, then R contains also the set I of integers.
- 4.10(a). The set I of integers is an integral domain, the unity of which is $1 \in N$.
- (b). The order of I coincides with the order of N .
- 4.11. The following theorems hold in N :
- (i) For every $a, b \in N$, $a < b + 1$ implies $a \leq b$.
 - (ii) For every $a, b, m, n \in N$, $(ab)^n = a^n b^n$, $a^m a^n = a^{m+n}$, and $(a^m)^n = a^{mn}$.
 - (iii) $1^n = 1$.
- 4.12. If $(x, y), (y, x), (z, z) \in J$, then
- (i) $(x, y) + (y, x) = 0$.
 - (ii) $(x, y) + (z, z) = (x, y)$.
 - (iii) $(x, y)(z, z) = (z, z)$.
- 4.13. The linear Diophantine equation $ax + by = n$, where $a, b, n, x, y \in I$, has a solution iff $(a, b) \mid n$.
- 4.14. If p is a prime and neither $p \mid a$ nor $p \mid b$, then $p \nmid ab$, and if $p \mid (a - b)$, then $p^2 \mid (a^p - b^p)$ and $p \mid ((a + b)^p - a^p - b^p)$.
- 4.15. If a and b are relatively prime, i.e. $(a, b) = 1$, then $(a + b, a - b) = 1$ or 2 .
- 4.16. For every $n \in N$, $a \equiv b \pmod{m}$ implies $a^n \equiv b^n \pmod{m}$.
- 4.17. If p is a prime and $a^2 \equiv b^2 \pmod{p}$, then $a \equiv b \pmod{p}$ or $a \equiv -b \pmod{p}$.
- 4.18. If p is a prime and $p \nmid a$, then $a^{p-1} \equiv 1 \pmod{p}$.
- 4.19. If none of $a_i, i = 1, 2, \dots, n$, is divisible by 2, then
- (i) $(a_1 a_2 \dots a_n - 1)/2 \equiv \sum_i (a_i - 1)/2 \pmod{2}$,
 - (ii) $((a_1 a_2 \dots a_n)^2 - 1)/8 \equiv \sum_i (a_i^2 - 1)/8 \pmod{8}$.
- 4.20. Given $28x \equiv 8 \pmod{44}$, find x .
- 4.21. Find x if $3x \equiv 1 \pmod{8}$ and $4x \equiv 3 \pmod{11}$.
- 4.22. Solve for x in the following simultaneous equations:
- $$x \equiv 1 \pmod{6}, \quad x \equiv 4 \pmod{9}, \quad x \equiv 7 \pmod{15}$$
- 4.23. If the characteristic of a field F is a prime p , then for every $a, b \in F$,
- (i) $(a + b)^p = a^p + b^p$, (ii) $(a - b)^p = a^p - b^p$, (iii) $(a_1 + a_2 + \dots + a_n)^p = a_1^p + a_2^p + \dots + a_n^p$.
- 4.24. Every field contains one, and only one, prime field in itself.
- 4.25. The characteristic of an ordered field is always zero.
- *4.26. Th. 4.1.2.5.19 holds conversely.

- *4.27. A polynomial in x_1, x_2, \dots, x_n , represented by the expression of Th. 4.1.2.5.19 vanishes iff $A_i = 0$, $i = 1, 2, \dots, m$.
- *4.28. $f(x_1, x_2, \dots, x_n) = g(x_1, x_2, \dots, x_n)$ in $D[x_1, x_2, \dots, x_n]$ iff the terms which constitute f coincide as a whole with those which constitute g .
- *4.29. If $f(x_1, x_2, \dots, x_n) g(x_1, x_2, \dots, x_n) = 0$ in $D[x_1, x_2, \dots, x_n]$, then $f(x_1, x_2, \dots, x_n) = 0$ or $g(x_1, x_2, \dots, x_n) = 0$.
- *4.30. $D[x_1, x_2, \dots, x_n]$ of Th. 4.1.2.5.19 is an integral domain.
- *4.31. In $D[x_1, x_2, \dots, x_n]$, $\deg(f(x_1, x_2, \dots, x_n)) = m$ and $\deg(g(x_1, x_2, \dots, x_n)) = n$ imply $\deg(f(x_1, x_2, \dots, x_n) g(x_1, x_2, \dots, x_n)) = m + n$.

4.32. If $i = \sqrt{-1}$, then

$$\bar{i} = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}, \quad \bar{j} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad \bar{k} = \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}$$

are quaternions.

4.33. By Df. 4.1.3.1.11 it follows that, for every $\bar{a}, \bar{b} \in \bar{Q}$,

- (i) $|\bar{a}| \geq 0$, and $|\bar{a}| = 0$ iff $\bar{a} = 0$.
- (ii) $|\bar{a} + \bar{b}| \leq |\bar{a}| + |\bar{b}|$.
- (iii) $|\bar{a} \cdot \bar{b}| = |\bar{a}| |\bar{b}|$.

4.34. Find vectors x, y, z which satisfy the following equations:

- (i) $2(-1, 3, 1) - 2x = (3, 4, 5)$.
- (ii) $3(2, 6, 1) + 3y = (6, 5, 4) - y$.
- (iii) $a + 2b + z = 4(z - a) + 5a - 4(z - b)$.

4.35. Determine whether the following vectors are linearly independent or not: $(1, 2, 7)$, $(-2, 5, 4)$, $(-1, 4, 5)$.

4.36. Any finite-dimensional vector space over a field F is isomorphic to $V_n(F)$.

4.37. If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and $f(x) = x^2 - (a+d)x + (ad - bc)$, then $f(A) = 0$.

4.38. Find a matrix X which satisfies the following equation: $\begin{bmatrix} 3 & 6 \\ 2 & 4 \end{bmatrix} X = [0]$.

4.39. Find a square matrix X such that $AX = XA$, where $A = \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix}$.

4.40. If A is nonsingular, then

- (i) $|A^{-1}| = 1/|A|$, (ii) $(A^{-1})^{-1} = A$, (iii) $(A^T)^{-1} = (A^{-1})^T$, (iv) $(A^{-1})^* = A/|A|$.

4.41. If A and B are two square matrices of the same order, then

$$|AB^T| = |B^T A| = |A^T B| = |BA^T| = |A| |B|$$

4.42. If A and B are two square matrices of the same order, then

- (i) $(A^T)^* = (A^*)^T$,
- (ii) $A^* B = BA^*$ if $AB = BA$.
- (iii) $(A^*)^* = A |A|^{n-2}$.
- (iv) $(A^*)^{-1} = A/|A|$.

4.43. Find $|A|, A^*, |A^*|, A^{-1}$, given $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 8 \\ 3 & 8 & 11 \end{bmatrix}$.

- 4.44. Find $([1] - A)([1] + A)^{-1}$, given $A = \begin{bmatrix} 0 & 1 & 2 \\ -1 & 0 & 3 \\ -2 & -3 & 0 \end{bmatrix}$.
- 4.45. If A is nonsingular and B is a matrix of the same order as A , there uniquely exist two matrices X and Y such that $AX = B$ and $YA = B$ and that $AB = BA$ implies $X = Y$.
- 4.46. If the sum of a finite number of the unity e in F is not zero, then the prime subfield of F is isomorphic to the set R of rational numbers.
- 4.47. If I is the ring of integers and E the ideal of even integers, then I/E is a ring.
- 4.48. The ideals $(10, 6)$ and $(6, 4)$ are principal ideals in I , generated by 2; i.e. $(10, 6) = (6, 4) = (2)$.
- 4.49. Prove, in terms of cyclic groups, that every ideal in the ring I of integers is principal.
- 4.50. If R is a commutative ring with unity and M is a maximal ideal in R , then M is a prime ideal.
- 4.51. Every Euclidean ring contains a unity element.
- 4.52. The integral domain $F(x)$ of polynomials in x over a field F is a Euclidean ring.
- 4.53. A field is a Euclidean ring.
- *4.54. If (a) and (b) are two nonzero principal ideals in an integral domain D , then $(a) = (b)$ iff a and b are associates in D .
- *4.55. If a and b are two nonzero elements in a Euclidean ring R without zero divisors, then either $\bar{n}(ab) = \bar{n}(a)$ or $\bar{n}(ab) > \bar{n}(a)$ according as b is a unit or not.
- *4.56. If a principal ideal ring R is also an integral domain and contains a prime element p in itself such that $p \mid ab$, where $a, b \in R$, then $p \mid a$ or $p \mid b$ or both.
- *4.57. If D is an integral domain in which factorization is uniquely feasible, then every prime element in D generates a prime ideal.
- *4.58. A ring with a nonzero identity is a sfield iff it has no proper right (or left) ideals.

Number Fields

§5.1.1 Rational Numbers

Df. 5.1.1.1 The *rational numbers* are the elements of the set R isomorphic to the field Q of quotients, defined by Th. 4.1.2.4.7; the set R as such is called the *rational number field*.

R is unique as the minimal field of all fields which contain the integral domain I of integers (cf. Th. 4.2.1.3-4 and Prob. 2-3 below); R as such contains no proper subfields which contain I . Hence the following alternative definition:

Df. 5.1.1.1a The rational number field R is the minimal field which contains the ring I of integers.

Stated otherwise, any field which contains I contains also R (cf. Prob. 4-5). R is further characterized by the following theorem:

Th. 5.1.1.2 R is a prime subfield of a field F , containing no subfield other than itself; i.e. R , in itself, is a prime field. (Cf. Prob. 5.)

Since the other prime field of F is the field P of integers modulo p , i.e. I_p (or $I/\{p\}$), as has been proved by Th. 4.2.1.4, every field is then a (simple) *extension* (cf. Df. 5.3.1.5) of either R or I_p .

Th. 5.1.1.3 There exists no isomorphic mapping other than automorphism in R , which is in fact of only one kind, viz. the *identity automorphism*, which maps every element of R into itself. (Cf. Prob. 6.)

Such a mapping may be considered an *order-automorphism* if R is ordered, as in the following theorem:

Th. 5.1.1.4 R is an ordered field (cf. Df. 4.1.2.4.11) iff $a/b > 0$ implies $ab > 0$, where $a, b \in I$ and $b \neq 0$. (Prob. 7-8.)

Since the quotient form may be replaced by the ordered-pair form of (a, b) or the inverse form of ab^{-1} , the theorem may be stated in terms of the latter: R is an ordered field iff $(a, b) > 0$ (or $ab^{-1} > 0$) implies $ab > 0$ for every $(a, b) \in R$ (or $ab^{-1} \in R$).

Th. 5.1.1.5 If a and b are any two distinct rational numbers, say $a < b$, then there exists a rational number c such that $a < c < b$. (Cf. Prob. 9.)

The set R , satisfying this theorem, is said to be *dense*, which is an important concept in analysis, where the density of a set of elements relative to an order relation plays a significant rôle. The order in R is also articulated in the following theorem:

Th. 5.1.1.6 R is an *Archimedean ordered field* (cf. Df. 4.1.2.4.12). (Cf. Prob. 10.)

In addition to the fundamental operative properties of the quotient field Q (cf. Th. 4.1.2.4.6), which of course hold for R , some other properties of R , which in turn hold for Q , are as follows:

Th. 5.1.1.7 For every $a, b, c, d, a', b' \in R$,

- (i) $0 < 1/a$ iff $0 < a$.
- (ii) $a > b$ and $c > 0$ imply $a/c > b/c$.
- (iii) $a > b > 0$ and $a' > b' > 0$ imply $a/b' > b/a'$.
- (iv) $a, b, c, d, a', b' > 0$ and $a/b > c/d$ imply $a/b > (a'a + b'b)/(a'b + b'd) > c/d$.
- (v) $a_i \in R, i = 1, 2, \dots, n$, implies $\sum_i a_i^2 \geq 0$.
- (vi) $a_i, b_i \in R, i = 1, 2, \dots, n$, implies $(\sum_i (a_i b_i))^2 \leq (\sum_i a_i^2)(\sum_i b_i^2)$. (Cf. Prob. 11-13.)

These properties hold for any ordered field \bar{F} , as a matter of fact, and Th. 5.1.1.7 will be carried into the field \bar{R} of real numbers, for instance, without any modification.

The following properties, which are the immediate results of Th. 5.1.1.4-5, 7, also hold for both R and \bar{R} :

- (vii) $a/b < c/d$ iff $abd^2 < b^2cd$.
- (viii) $0 < 1/b < 1/a$ if $0 < a < b$.
- (ix) $0 > 1/a > 1/b$ if $0 > b > a$.

There are, of course, many other similar and deducible properties (cf. Prob. 12).

Th. 5.1.1.8 For every $a, b, c, d, a', b', c', d' \in R$, $a/b = c/d$ and $a'/b' = c'/d'$ imply

- (i) $(ab' + a'b)/bb' = (cd' + c'd)/dd'$,
- (ii) $aa'/bb' = cc'/dd'$

(Cf. Prob. 15.)

Solved Problems

1. A complex R' of the set R of rational numbers is isomorphic to the set I of integers iff there exists a 1-1 mapping f of R' into I such that, for any $x, y \in I$,

- (i) $f(x + y) = f(x) + f(y)$,
- (ii) $f(xy) = f(x)f(y)$,
- (iii) $x > y$ implies $f(x) > f(y)$.

PROOF:

Let every element of R' be of the form $1/a$, where $a \in I$; then any element p/q , $p, q \in I$, of R will be contained in R' iff $q/p \in I$, since $1/(p/q) = q/p$.

Now let the mapping f be defined by

$$f: f(a) = 1/a, \text{ for every } a \in I$$

Then $x \in I \Leftrightarrow x/1 = f(x) \in R'$ and $y \in I \Leftrightarrow y/1 = f(y) \in R'$ imply

- (i) $x + y \in I \Leftrightarrow f(x + y) \Leftrightarrow f(x) + f(y) = (x + y)/1 \in R'$,
- (ii) $xy \in I \Leftrightarrow f(xy) \Leftrightarrow f(x)f(y) = xy/1 \in R'$,
- (iii) $x > y \Leftrightarrow f(x) > f(y) \Leftrightarrow x/1 > y/1$,

which completes the proof.

Note. f in this context is evidently an order-isomorphism (cf. Th. 4.1.2.3.8).

2. A field F containing the set I of integers is a minimal field, in fact, the set R of rational numbers itself, iff every element of F is of the form p/q , $p, q \in I$.

PROOF:

If F contains I , and if every element of F is of the quotient form, then, by Th. 4.2.1.3-5, F is a minimal field Q , coinciding with R .

Conversely, assume F to be a minimal field. Then any field A , every element of which is of the quotient form, satisfies the following properties of Q or R (cf. Th. 4.1.2.4.6): given $b, d \neq 0$,

$$(i) \quad a/b = c/d \text{ iff } ad = bc,$$

$$(ii) \quad (a/b) \pm (c/d) = (ad \pm bc)/bd,$$

$$(iii) \quad (a/b)(c/d) = ac/bd,$$

$$(iv) \quad (a/b)/(c/d) = ad/bc.$$

Hence A is a subfield of Q or R and also, by Prob. 1, contains I , i.e. $A = R$, completing the proof.

3. Any minimal field which contains I is isomorphic to Q or R ; i.e. the rational number field R is unique except for its isomorphisms.

PROOF:

Let A and A' be two minimal fields containing I individually; then, by Prob. 2 above, all elements of A and A' are of the quotient form. Now, if a mapping of A into A' is defined to be

$$f: f(x) = y$$

for every $x = a/b$ and $y = a'/b'$, where the choice of a and b is arbitrary, then, by the property (ii) of Prob. 2 above,

$$(i) \quad f(x) = a/b = y \text{ and } f(x') = c/d = y' \text{ imply } f(x) + f(x') \leftrightarrow y + y'$$

Likewise, by the property (iii) of Prob. 2,

$$f(x) \cdot f(x') \leftrightarrow y \cdot y'$$

Hence f is an isomorphism, completing the proof.

4. Any field A which contains the set I of integers contains the set R of rational numbers.

PROOF:

Since the meet of any number of subfields of a field F is again a subfield of F (cf. §4.2.1.1, Prob. 4), the meet of all subfields of A containing I is a subfield containing I , which in turn is contained in any subfield containing I . Hence A must be a minimal subfield, thus isomorphic to R , by Prob. 3.

5. Any field A of characteristic zero contains one, and only one, subfield B , which in turn is isomorphic to the rational number field R .

PROOF:

Since the characteristic of A is zero, any integer $a \neq 0$ implies $ae \neq 0$ for $e \in A$; also $b \in I$ and $a \neq b$ imply $a - b \neq 0$ and $ae - be = (a - b)e \neq 0$. Hence the mapping f of I into A , $a \leftrightarrow ae$, is 1-1; moreover, it is isomorphic, since $x \leftrightarrow xe$ and $y \leftrightarrow ye$ for every $x, y \in I$ imply

$$x + y \leftrightarrow f(x) + f(y) \leftrightarrow xe + ye = (x + y)e$$

and

$$x \cdot y \leftrightarrow f(x) \cdot f(y) \leftrightarrow (xe)(ye) = (xy)e \quad (1)$$

Likewise, for any rational numbers $x, y \in B$, where $x = a/b$ and $y = c/d$,

$$x + y \leftrightarrow f(x) + f(y) \leftrightarrow xe + ye = (x + y)e$$

and

$$x \cdot y \leftrightarrow f(x) \cdot f(y) \leftrightarrow (xe)(ye) = (xy)e \quad (2)$$

$$\text{since } xe + ye = (ae/be) + (ce/de) = ((ae)(de) + (be)(ce))/((be)(de)) = (ad + bc)e/(bd)e = (x + y)e$$

by (ii) in Prob. 2 above, and

$$(xe)(ye) = (ae/be)(ce/de) = ((ae)(ce))/((be)(de)) = (ac)e/(bd)e = (xy)e$$

by (iii) in Prob. 2. Hence, by (1) and (2), A contains B , which is isomorphic to R .

Furthermore, if a subfield B' is contained in A and also isomorphic to R , $B' = B$, because there exists no isomorphism in R other than the unique identity automorphism, which is the following theorem (Prob. 6 below).

6. Prove Th. 5.1.1.3.

PROOF:

Let f be the mapping of R into R , hence an automorphism $f: a \leftrightarrow a'$, where $a, a' \in R$. It follows, then, that $a' \cdot e' = a'$ for every a' , since $a \cdot e = a$ for every a . But, by F9, the unity is unique, which implies $e' = e$ (or $1' = 1$).

If a is an integer, then

$$a' = 1' + 1' + \dots + 1' = 1 + 1 + \dots + 1 = a$$

which means that the automorphism $a \leftrightarrow a'$ implies $a' = a$. Likewise $(a')^{-1} = (a^{-1})'$ since $a'(a^{-1})' = 1' = 1 = aa^{-1}$, and also $0' = 0$, and $-a' = (-a)'$, verifying that $a \leftrightarrow a'$ implies $a' = a$ for every integer a .

If, more generally, a is a rational number of the quotient form, say $a = p/q$, $p, q \in I$, and if $f: p/q \leftrightarrow (p/q)'$ is an automorphism, then

$$a' = (p/q)' = p'/q' = p/q = a$$

since it has already been established that $p \leftrightarrow p'$ and $q \leftrightarrow q'$ imply $p = p'$ and $q = q'$ for every $p, q \in I$.

Hence it follows here again that $a \leftrightarrow a'$ implies $a = a'$, verifying that R has no automorphism but the identity automorphism.

7. Prove Th. 5.1.1.4.

PROOF:

If $a/b = c/d$ for every $a, b, c, d \in I$ and $b, d \neq 0$, then $ad = bc$, by Prob. 2, (i), which implies $abd^2 = b^2cd$, where obviously $b^2, d^2 > 0$ (cf. Th. 4.1.2.2.9), which in turn implies that cd has the same sign as b^2cd just as ab has the same sign as abd^2 . Hence $a/b > 0$ and $a/b = c/d$ imply $c/d > 0$, verifying that equals of positive elements are positive.

Conversely, Df. 4.1.2.2.5 is satisfied in this context, since for every $x = a/b$ and $y = c/d$, where a, b, c, d are defined as above,

- (i) $x > 0$ and $y > 0$ imply $x + y > 0$; for $x + y = (ad + bc)/bd > 0$ since $(ad + bc)bd = (ab)d^2 + (cd)b^2 > 0$,
- (ii) $x > 0$ and $y > 0$ imply $xy > 0$; for $xy = ac/bd > 0$ since $(ac)(bc) = (ab)(cd) > 0$, and
- (iii) $x > 0$ excludes the alternatives of $x = 0$ and $x < 0$.

Furthermore, by Prob. 1 above, the order of positive fractions coincides with that of special fractions of the form $a/1$, which in fact represent integers, where $a/1 > 0$ iff $a \cdot 1 > 0$, completing the proof.

8. The ordering in Th. 5.1.1.4 is unique.

PROOF:

Assume that there exists the second mode of ordering, which does not change the sense of positiveness. Then $x = a/b > 0$ iff $ab > 0$, since $a/b > 0$ implies $b^2(a/b) > 0$, i.e. $ab > 0$, and conversely, $ab > 0$ implies $a/b > 0$; for, otherwise, $-(a/b) \cong 0$, i.e. $b^2(-(a/b)) \cong 0$, which implies $-ab \cong 0$, contradictory to the assumption. Hence the theorem must follow.

9. Prove Th. 5.1.1.5.

PROOF:

Since $a < b$, it follows at once that $a + a < a + b$, i.e. $2a < a + b$, which implies $a < (a + b)/2$. On the other hand, $a < b$ implies $a + b < b + b$, i.e. $a + b < 2b$, which implies $(a + b)/2 < b$. Hence, combining the results, $a < (a + b)/2 < b$, which implies the existence of $c = (a + b)/2$ such that $a < c < b$.

10. If $x = a/b$ and $y = c/d$, $a, b, c, d \in N$, then there exists n , $n \in N$, such that $nx > y$.

PROOF:

By Prob. 2, (i), $n(a/b) > c/d$ iff $n(ad) > bc$, which is always the case, however, since n can be always equated to $2bc$. Then, by hypothesis, $ad \cong 1$ and $n(ad) = 2bc(ad) = 2ad(bc) > bc$, i.e. $nx > y$, completing the proof.

11. For every $a, b, c, d, a', b' \in R$,

- (i) $0 < 1/c$ iff $0 < c$.
- (ii) $a > b$ and $c > 0$ imply $a/c > b/c$.
- (iii) $a > b > 0$ and $a' > b' > 0$ imply $a/b' > b/a'$.

PROOF:

- (i) This is merely a special case of Th. 5.1.1.4 where substitution is carried out as $a=1$ and $b=c$.
- (ii) Since $c > 0$ implies $1/c > 0$, by (i) above, and since $a-b > 0$, by hypothesis, it follows from Th. 5.1.1.4 that $(a-b)(1/c) > 0$, i.e. $(a/c) - (b/c) > 0$ or what is the same: $a/c > b/c$, which was to be proved.
- (iii) Since $aa' > bb' > 0$ and $a'b' > 0$, by hypothesis, it follows from (i) above that $aa'/a'b' > bb'/a'b'$, i.e. $a/b' > b/a'$.

Second Proof. Since $a' - b' > 0$ and $a'b' > 0$, by hypothesis, which implies $1/a'b' > 0$ as in (i) above, it follows from Th. 5.1.1.4 that $(a'-b')/a'b' = (1/b') - (1/a') > 0$, i.e. $1/b' > 1/a'$, which implies $a(1/b') > b(1/a')$, since $a > b$, by hypothesis. Hence $a/b' > b/a'$.

12. If $a, b, c, d, a', b' \in N$ and $a/b > c/d$, then

$$a/b > (a'a + b'c)/(a'b + b'd) > c/d$$

PROOF:

Since $a/b < c/d$, $b, d > 0$, by hypothesis, it follows that $ad - bc < 0$. Hence $a', b' > 0$ implies

- (i) $(ad - bc)a' = ada' - bca' = ada' + cdb' - cdb' - bca'$
 $= d(aa' + cb') - c(ba' + db') > 0$, i.e. $(a'a + b'c)/(a'b + b'd) > c/d$, and
- (ii) $(ad - bc)b' = adb' - bcb' = adb' + aba' - aba' - bcb'$
 $= a(ba' + db') - b(aa' + cb') > 0$, i.e. $a/b > (a'a + b'c)/(a'b + b'd)$.

Hence, combining the results of (i) and (ii),

$$a/b > (a'a + b'c)/(a'b + b'd) > c/d$$

Note that this is merely an explicit form of Prob. 9 (Th. 5.1.1.5).

13. For every $a_i, b_i \in R$, $i = 1, 2, \dots, n$,

- (i) $\sum_i a_i^2 \geq 0$,
- (ii) $(\sum_i (a_i b_i))^2 \leq (\sum_i a_i^2)(\sum_i b_i^2)$.

PROOF:

- (i) Let $a_i = x_i/y_i$, where $x_i, y_i \in I$ and $y_i \neq 0$; then $(a_i)^2 = (x_i)^2/(y_i)^2 > 0$ since $x_i^2, y_i^2 > 0$ by Th. 4.1.2.2.9, and $a_i^2 = 0$ iff $x_i = 0$. Hence

$$\sum_i a_i^2 \geq 0, \quad i = 1, 2, \dots, n$$

$$\begin{aligned} \text{(ii)} \quad & (\sum_i a_i^2)(\sum_i b_i^2) - (1/2) \sum_i \sum_j (a_i b_j - a_j b_i)^2 \quad i, j = 1, 2, \dots, n \\ &= (\sum_i a_i^2)(\sum_i b_i^2) - (\sum_i \sum_j a_i^2 b_j^2)/2 - (\sum_i \sum_j a_j^2 b_i^2)/2 + \sum_i \sum_j (a_i b_j a_j b_i)^2 \\ &= (\sum_i a_i^2)(\sum_i b_i^2) - (\sum_i a_i^2 \sum_i b_i^2)/2 - (\sum_i a_i^2 \sum_i b_i^2)/2 + \sum_i (a_i b_i) \sum_j (a_j b_j) \\ &= (\sum_i (a_i b_i))(\sum_i (a_i b_i)) = (\sum_i (a_i b_i))^2, \end{aligned}$$

$$\text{i.e.} \quad (\sum_i (a_i b_i))^2 = (\sum_i a_i^2)(\sum_i b_i^2) - (1/2) \sum_i \sum_j (a_i b_j - a_j b_i)^2 \quad \text{viz.} \quad (\sum_i (a_i b_i))^2 \leq (\sum_i a_i^2)(\sum_i b_i^2).$$

Note. (ii) is in fact the generalized form of the (Cauchy-)Schwarz inequality (sometimes called Buniakovsky's inequality), which can be employed for the set \bar{R} of real numbers without any modification.

14. For every $a, b, c \in I$, $(a/b) + (a/c) = a/(b+c)$ implies that $a=0$ or $b^2 + bc + c^2 = 0$, and that $a=0$ in an ordered field.

PROOF:

- (i) Since $(a/b) + (a/c) = (ac+ab)/bc$, it follows from hypothesis that $(ac+ab)/bc = a/(b+c)$ or $(ac+ab)(b+c) = abc$, by Prob. 2, (i) above, which implies $a(b^2 + bc + c^2) = 0$.

Hence the necessary and sufficient condition for the identity at issue is that $a=0$ or $b^2 + bc + c^2 = 0$.

- (ii) If the identity is in an ordered field, however, then $b^2 + bc + c^2 \neq 0$, since $b, c \neq 0$, by hypothesis, which implies $b^2 + bc + c^2 = (b+c)^2 - bc > 0$ ($\because (b+c)^2 > 2bc > bc$ since $(b+c)^2 - 2bc = b^2 + c^2 > 0$). Hence the alternative: $a=0$, completing the proof.

15. Prove Th. 5.1.1.8.

PROOF:

- (i) Since $a/b = c/d$ and $a'/b' = c'/d'$, i.e. $ad = bc$ and $a'd' = b'c'$, by hypothesis, it follows that

$$\begin{aligned}(ab' + a'b)dd' &= (ab')(dd') + (a'b)(dd') = (ad)(b'd') + (bd)(a'd') \\ &= (bc)(b'd') + (bd)(b'c') = (cd' + c'd)bb'\end{aligned}$$

Hence $(ab' + a'b)/bb' = (cd' + c'd)/dd'$.

- (ii) Likewise, $(aa')(dd') = (ad)(a'd') = (bc)(b'c') = (bb')(cc')$.

Hence $aa'/bb' = cc'/dd'$.

§5.1.2 Real Numbers

Df. 5.1.2.1 A proper complex S of the set R of rational numbers, viz. $S \subset R$, and $S \neq R$, is called a *Dedekind cut* (or *D-cut*, or more simply, *cut*),

- (i) if $s \in S$ and $r \in R$ such that $r < s$ imply $r \in S$, and
(ii) if $s \in S$ implies $s' > s$ for some $s' \in S$.

Example:

Any nonempty subset T of R , defined by $T = \{t \mid t \in R < r \in R\}$, is a *D-cut*, called a *D-cut at r* ; T as such may be denoted by T_r . T_3 , then, is a cut at 3, designating the set of all rational numbers less than 3, which may be pictorially represented by a segment T in Fig. 5.1.2.a below:



Fig. 5.1.2a

The other part of the line, denoted by T' above, is evidently the complement (cf. Df. 2.3.3) of T , designating the set of all rational numbers which are not in T , viz. $T' = \{t \mid t \in R \geq 3 \in R\}$. T' as such may be denoted by T'_3 in this context, and in general, $T' = \{t \mid t \in R \geq r \in R\}$.

It is also pictorially evident in the same context that, given a cut T , some members in T must be quite close to some numbers in T' . This fact, intuitively obvious, is formulated in the following theorem:

Th. 5.1.2.2 There exists an element s of a cut S such that $s + r \in S'$ for any positive rational number r . (Cf. Prob. 1.)

Since a cut is a set, a collection of cuts is a class, denoted by C , in which addition and multiplication are defined as follows:

Df. 5.1.2.3 For every $A, B \subset C$, and for every $a \in A$ and $b \in B$,

- (i) $A \cup B = \{a + b\}$ for every $a \in A, b \in B$;
- (ii) $A \cap B = \{a \cdot b\}$ for every $a, b > 0$ or $a, b < 0$;
 $= \{-a \cdot b\}$ for every $a > 0$ and $b < 0$ or $a < 0$ and $b > 0$;
 $= \{0\}$ for every $a = 0$ or $b = 0$.

Note. $\{0\}$, as will be seen below, is a cut at 0, and as such may be denoted by 0^* or even more simply by 0 if no confusion is to result.

Th. 5.1.2.4 The join (or sum, as is often called) of two cuts A and B of the class C , defined as above, uniquely defines a cut which belongs to C itself; so does the meet (or product) of A and B . (Cf. Prob. 2, 12.)

This is evidently the closure property of binary operations in C , from which an algebra of cuts as a special algebra of sets (cf. §2.3.1-3) can be developed, verifying associativity, commutativity, etc. As a matter of fact, these joins and meets of cuts satisfy all the properties of F1-11, proving C to be a field (cf. Prob. 18), which is also ordered, as is exemplified in the following theorems:

Th. 5.1.2.5 Given two members A and B of C , A is less than B , i.e. $A < B$, iff A is a proper subset of B , i.e. $A \subset B$ and $A \neq B$. (Cf. Prob. 9.)

Transitivity follows at once, as can be readily verified (cf. Prob. 11), as well as trichotomy:

Th. 5.1.2.6 Given any $A, B \subset C$, one and only one of the following three cases holds: (i) $A < B$, (ii) $A = B$, (iii) $A > B$. (Cf. Prob. 12.)

Once order is introduced in C , a subclass of C is first verified to be order-isomorphic to the rational number field R on the strength of the following theorem and definition:

Th. 5.1.2.7 Given any $A, B \subset C$, where $A < B$ (or dually, $A > B$), there always exists a cut C in C such that $A < C < B$ (or $A > C > B$). (Cf. Prob. 20.)

Df. 5.1.2.8 The third cut C in Th. 5.1.2.7 is called a *rational cut*, and if a rational number, say r , is explicitly assigned to the cut, it is then denoted by C_r or more simply by r^* .

Example:

C_0 or 0^* designates the cut at 0.

The class C^* of rational cuts, however, cannot exhaust the class C ; i.e. there exist cuts other than the cuts at rational numbers. There are many, in fact infinitely many, and indeed *uncountably* (cf. Df. 2.1.13) infinitely many gaps between rational cuts, even though Th. 5.1.2.5 allows the cutting at rational numbers to go on *ad infinitum*. Infinite as it may be, the whole set of rational numbers is still *countable* (cf. §2.1, Prob. 12).

The cuts which are not rational do exist, individually exemplified by such irrational numbers as $\sqrt{2}$ (cf. §2.1, Prob. 4 note); such cuts, then, may be called *irrational cuts*, the existence of which in general is assured by the following theorem:

Th. 5.1.2.9 If A and B are two mutually exclusive sets of rational numbers (or what is the same, two disjoint classes of rational cuts) such that every rational number is either in A or in B , but never in both, and $a < b$ for every $a \in A$ and $b \in B$, then there exists either a *leap* whenever A has the largest number and B has the smallest or a *gap* whenever neither A has the largest number nor B has the smallest. (Cf. Prob. 23-25.)

Between rational numbers (or cuts), therefore, there always exist *discontinuities*, either as leaps or as gaps, which are then to be filled by irrational numbers (or cuts). The class C^* of rational cuts is thus a subclass of the class C of cuts, which contain both rational and irrational cuts, yielding the real number field as follows:

Df. 5.1.2.10 An ordered field isomorphic to the class C of cuts is called the *real number field*, denoted by \bar{R} (or R^*), each element of which is called a *real number*.

In terms of rational numbers, a real number is further characterized by the following theorem:

Th. 5.1.2.11 Every real number is the l.u.b. (or sup. (cf. Df. 2.4.1.6)) of a set of rational numbers. (Cf. Prob. 28.)

Example:

$\sqrt{2} = \text{l.u.b. } (1.4, 1.41, 1.414, 1.4142, \dots)$, which may be considered also in terms of g.l.b. (or inf. (cf. Df. 2.4.1.7)), viz.

$$\sqrt{2} = \text{g.l.b. } (1.5, 1.43, 1.415, 1.4143, \dots)$$

which yields the dual of Th. 5.1.2.11, viz.

Th. 5.1.2.11a Every real number is the g.l.b. (or inf.) of a set of rational numbers. (Cf. Prob. 28.)

In the same fashion, then, the D -cut in general may be defined also in terms of lower and upper bounds, viz.

Df. 5.1.2.12 A (Dedekind) cut S in any ordered field \bar{F} uniquely determines a pair of proper complexes L and U of \bar{F} such that

- (i) L is the set of all lower bounds to every $u \in U$, and
- (ii) U is the set of all upper bounds to every $l \in L$.

Since Th. 5.1.1.5 or its equivalent, e.g. Th. 5.1.2.7, can be readily generalized to the existence of a rational number c or a real number d such that $a < c < b$ or $a < d < b$ for every $a, b \in \bar{R}$ (cf. Prob. 26-27), Th. 5.1.2.11 is also generalized as follows:

Th. 5.1.2.13 If A is a set of real numbers bounded from above, then there uniquely exists a l.u.b. s such that

- (i) $a \leq s$ for every $a \in A$, and
- (ii) there exist some numbers, say a' , in A such that $a - p < a'$, where p is any positive number. (Cf. Prob. 29.)

As is but natural in view of its relation to Th. 5.1.2.11, this theorem also has its dual:

Th. 5.1.2.13a If A is a set of real numbers bounded from below, then there uniquely exists a g.l.b. t such that

- (i) $b \leq t$ for every $a \in A$, and
- (ii) there exist some numbers, say a' , in A such that $a' < a + p$, where p is any positive number. (Cf. Prob. 29.)

\bar{R} as such is also Archimedean ordered, viz.

Th. 5.1.2.14 If $a, b \in \bar{R}^+$ (i.e. $a, b \in \bar{R}$ and $a, b > 0$), then there exists $n \in I^+$ such that $na > b$. (Cf. Prob. 30.)

The Archimedean ordered field \bar{R} is further characterized by its completeness (cf. Prob. 42), in the sense of the following definition:

Df. 5.1.2.15 (Axiom of Completeness). An ordered field F is called *complete* if every complex of F having an upper bound has a l.u.b. and every complex of F having a lower bound has a g.l.b.

Hence \bar{R} is now known to be an ordered and complete field; in fact, \bar{R} is the only complete ordered field (to which, of course, some fields may be isomorphic, cf. Prob. 44-45).

* * * * *

As has already been explicit in Th. 5.1.2.13-14, the set \bar{R} of real numbers may be defined in terms of *Cauchy-Cantor sequences*, the development of which may be considered more analytic and less algebraic than that of Dedekind cuts. It begins with infinite sequences in general.

Df. 5.1.2.16 An *infinite sequence* is a (single-valued) function: $f(n) = a_n$, where $a_n \in F$, the domain (cf. Df. 2.2.2.6) of which is the set N of natural numbers.

The function f is then prescribed by the correspondence of a unique value or *term* of the sequence to a positive integer, viz.

$$1 \leftrightarrow a_1, 2 \leftrightarrow a_2, \dots, n \leftrightarrow a_n, \dots,$$

where the n th term a_n is sometimes called the *general term* of the infinite sequence, which itself is then represented by $\{a_n\}$. Unless stated otherwise, sequences will mean infinite sequences in the following pages.

Df. 5.1.2.17 A *subsequence* of a sequence $S = \{a_n\}$ is a sequence A , denoted by $\{a_{n_k}\}$, representing a_{n_1}, a_{n_2}, \dots , where $n_1 < n_2 < \dots$ in the original order of S .

Example:

$1/3, 1/5, \dots$ is a subsequence of a sequence $1/2, 1/3, 1/4, 1/5, 1/6, \dots$

Df. 5.1.2.18 The sequence $S = \{a_n\}$ is said to have the *limit* a , denoted by $\lim_{n \rightarrow \infty} a_n = a$, or more simply $\lim a_n = a$, iff corresponding to any natural number p there exists a natural number $m = m(p)$ such that $|a_n - a| < p$ whenever $m < n$. If S has the limit a , then it is said to be *convergent* (to a); otherwise it is *divergent*.

Example:

$1, 1/2, \dots, 1/n, \dots$ converges to 0, since $|a_n - a| = |1/n - 0| = 1/n$, and since $1/n < p$ whenever $1/p < n$; i.e. $m(p) = 1/p$ in this example.

As can be readily verified (cf. Prob. 31-33), the alteration of a finite number of terms of a sequence S has no effect on convergence or divergence or limit, which is always unique if S converges, and any subsequence of a convergent sequence S converges to the same limit of S .

Df. 5.1.2.19 A sequence $S = \{a_n\}$ is *bounded* iff there exists a positive number b such that $|a_n| < b$ for any $n \in N$.

Stated otherwise, S is bounded iff all of its terms are contained in some finite interval; e.g. any convergent sequence is bounded (cf. Prob. 34).

Df. 5.1.2.20 If $\{a_n\}$ and $\{b_n\}$ are two sequences, the sequences $\{a_n + b_n\}$, $\{a_n - b_n\}$, and $\{a_n b_n\}$ are called their *sum*, *difference*, and *product*, respectively; if furthermore $b_n \neq 0$, the sequence $\{a_n/b_n\}$ is their *quotient*. In particular, sums and products may be extended to any finite number of sequences.

As can be verified without difficulty (cf. Prob. 36-40), the sum (or difference or product or quotient) of two convergent sequences is again a convergent sequence, its limit being the sum (or difference or product or quotient) of the respective limits.

Df. 5.1.2.21 (Cauchy's Convergence Condition). A sequence $S = \{a_n\}$, $a_n \in F$, is a *Cauchy* (-Cantor) *sequence* (or simply *C-sequence*) if corresponding to a positive number $p \in F$ there exists a natural number $m = m(p)$ such that $|a_u - a_v| < p$ whenever $m < u$ and $m < v$.

Example:

Given a sequence:

$$s_n = 1 - (1/2) + (1/3) - (1/4) + \dots + (-1)^{n-1}/n + \dots$$

$u > v$ implies

$$|s_u - s_v| = 1/(v+1) - (1/(v+2) - 1/(v+3)) - \dots < 1/(v+1)$$

Hence, if m is so chosen that $m+1 > 1/p$,

$$|s_u - s_v| < 1/(m+1) < p \quad \text{when } u, v > m$$

The sequence thus converges, obviously to a number between 1 and $1/2$, since

$$s_n = (1 - 1/2) + (1/3 - 1/4) + \dots = 1 - (1/2 - 1/3) - (1/4 - 1/5) - \dots$$

The field F , to which any of a_n belongs, is still the rational number field R (or its isomorphs), since the real number field \bar{R} has not been defined in this context; once defined, however, Df. 5.1.2.21 will be stated in terms of \bar{R} without any modification except that $a_n \in \bar{R}$.

Th. 5.1.2.22 Any convergent sequence in F is a *C-sequence*. (Cf. Prob. 41.)

This theorem and others (cf. Prob. 42) lead to the following theorem and definition:

Th. 5.1.2.23 An ordered field F which contains R is Archimedean ordered iff every element of F is the limit of a sequence $\{a_n\}$, where $a_n \in R$. (Cf. Prob. 43.)

The real number field \bar{R} is thus defined in terms of sequences.

Df. 5.1.2.24 The complete and Archimedean ordered field is called a *continuum*, and the real number field is a continuum \bar{R} which contains as a subfield the rational number field R .

It must be remembered that the continuum \bar{R} has a cardinal number $o(\bar{R}) = c$ (or the so-called "aleph one", cf. Df. 2.1.16).

Solved Problems

1. Prove Th. 5.1.2.2.

PROOF:

Since $r \in S'$ implies that $0+r \in S'$ ($\because 0 \in A$, as A has already been defined to be non-vacuous), the theorem is trivial if $r \in S'$.

Let $r \notin S'$; then $r \in S$, since $S \cup S' = R$ (cf. Df. 2.3.3). If a is a fixed element of S' , then, by Th. 5.1.1.6, there exists $n \in N$ such that $nr > a$, which implies $nr \in S'$, since $nr \notin S'$ would imply $nr \in S$ and consequently $a \in S$, contradictory to the present assumption.

Now, since there exists at least one element, viz. n , the subset of N which contains n is not empty; also, since $n > 1$ ($\because n=1$ implies $1 \cdot r = r \in S'$, contradictory to the assumption), it follows that $(n-1)r \in S$, which implies $(n-1)r + r \in S'$, i.e. $s = (n-1)r$, completing the proof.

Note. This theorem holds also conversely (cf. Prob. 2 below).

2. The join of two cuts A and B of C is again a cut in C .

PROOF:

Let $A \cup B = C$; then C is a proper complex of R , since $A \neq 0$ and $B \neq \emptyset$ imply $A \cup B = C \neq \emptyset$, and since $A \neq R$ and $B \neq R$ imply that there exist $a', b' \in R$ such that $a' \notin A$ and $b' \notin B$, which in turn imply $a' + b' \notin C$, i.e. $C \neq R$. Furthermore,

- (i) If $r \in R$ and $a + b = c$, where $a \in A$, $b \in B$, $c \in C$, such that $r < c$, then $r \in C$, since $r < c = a + b$, i.e. $r - a < b$, which implies $r - a \in B$, i.e. $r = a + (r - a)$ and consequently $r \in A \cup B = C$.
- (ii) If $a + b = c \in C$, where $a \in A$ and $b \in B$, then there exists $d \in A$ such that $d > a$, since A is a cut (cf. Df. 5.1.2.1, ii); this implies $d + b > a + b = c$ and $d + b = c' \in C$, i.e. $c' > c$ and $c' \in C$.

Df. 5.1.2.1 is thus satisfied with respect to C ; hence $C = A \cup B$ is again a cut, completing the proof.

Note. As is explicit in (ii) of the proof, the condition (ii) of Df. 5.1.2.1 can be more simply stated: S contains no largest rational.

3. If C_x is a cut at x , where $x \in R$, then

$$C_p \cup C_q = C_{p+q}$$

where $p, q \in R$ (cf. Df. 5.1.2.8).

PROOF:

If $a + b \in C_p + C_q$, where $a \in C_p$ and $b \in C_q$, then $a < p$ and $b < q$, which imply $a + b < p + q$.

Hence $a + b \in C_{p+q}$, i.e. $C_p \cup C_q \subseteq C_{p+q}$.

Conversely, if $c \in C_{p+q}$ such that $c < p + q$, then $(p + q) - c = d$, where obviously $d \in R^+$. Hence $c = (p + q) - d = (p - (d/2)) + (q - (d/2))$, where $p - (d/2) < p$ and $q - (d/2) < q$, which imply $p - (d/2) \in C_p$ and $q - (d/2) \in C_q$, i.e. $c \in C_p + C_q$, which implies $C_{p+q} \subseteq C_p \cup C_q$.

Combining both results, $C_p \cup C_q = C_{p+q}$.

4. Addition in C is associative and commutative.

PROOF:

If $C_a, C_b, C_c \subset C$, then by Prob. 3,

$$(i) \quad C_a \cup (C_b \cup C_c) = C_a \cup C_{b+c} = C_{a+(b+c)} = C_{(a+b)+c} = C_{a+b} \cup C_c = (C_a \cup C_b) \cup C_c;$$

$$(ii) \quad C_a \cup C_b = C_{a+b} = C_{b+a} = C_b \cup C_a.$$

5. If $A, B \subset C$ and $A < B$, then $A \cup C < B \cup C$ for every $C \subset C$.

PROOF:

Since $A < B$, there exists $a \in R$ such that $a \in B$ and $a \notin A$. Choose, then, $b \in R$ such that $a < b$ and $b \in B$; then, by Prob. 5, there exist $c, d \in R$ such that $d - c = b - a$, i.e. $b + c = a + d$, where $c \in C$ and $d \notin C$, which implies

$$b + c \in B \cup C \tag{1}$$

On the other hand, $a \notin A$ and $d \notin C$ imply that $a + d = b + c > e + f$ for every $e \in A$ and $f \in C$, which implies

$$b + c \notin A \cup C \quad (2)$$

Hence, by (1) and (2), $A \cup C < B \cup C$.

Note. Th. 5.1.2.5 is taken for granted here, which is to be proved in a different context (cf. Prob. 9); within the frame of this proof, then, Th. 5.1.2.5 may be considered a definition, without falling into a vicious circle.

6. If $A \subset C$, then there uniquely exists $B \subset C$ such that $A \cup B = 0^*$.

PROOF:

Since $A \cup B_1 = A \cup B_2$, where $B_1 \neq B_2$, is a direct contradiction to Prob. 5, there is at most one such B to be considered in the following.

Let B be the set of all rational numbers a such that $-a$ is an upper bound of A , but not the smallest number. Then B is a cut, since $B \neq \emptyset$, $B \neq R$, and

- (i) $a \in B$ and $b < a$, where $b \in R$, imply $-a \in A$ and $-b > -a$, so that $-b$ is an upper number of A , but not the smallest, which implies $b \in B$, satisfying Df. 5.1.2.1, (i); and
- (ii) $a \in B$ implies that $-a$ is an upper number of A , but not the smallest, so that there exists $c \in R$ such that $-c < -a$ and $-c \notin A$, which imply $-b < -d < -a$ (cf. Th. 5.1.1.5), where $d = (a + c)/2$, such that $-d$ is an upper number of A , but not the smallest; hence $d > a$ and $d \in B$, satisfying Df. 5.1.2.1, (ii).

Furthermore, if $p \in A \cup B$, then there exist $q \in A$ and $r \in B$ such that $p = q + r$, which implies $-r \notin A$, $-r > q$, $q + r < 0$, and $p \in 0^*$.

Conversely, if $p \in 0^*$, then $p < 0$ and, by Prob. 4, there exist $q \in A$ and $r \notin A$, where r is not the smallest upper number of A , such that $r - q = -p$, which implies, since $-r \in B$, that

$$p = q - r = q + (-r) \in A \cup B$$

Hence $A \cup B = 0^*$.

Note. Notationally, B may be written as $-A$ in this context, i.e. $A \cup B = A \cup (-A) = 0^*$. It should be emphasized that $-A$ cannot be replaced by A' (the complement of A ; cf. Fig. 5.1.2.a) in this specific algebra of sets.

7. If $A, B \subset C$, then there uniquely exists $C \subset C$ such that $A \cup C = B$.

PROOF:

As in Prob. 6, it follows directly from Prob. 5 that there exists at most one such C , since $C_1 \neq C_2$ implies $A \cup C_1 \neq A \cup C_2$.

Let $C = B \cup (-A)$ (cf. Prob. 6 note above); then, by Prob. 4,

$$A \cup C = A \cup (B \cup (-A)) = A \cup ((-A) \cup B) = (A \cup (-A)) \cup B = 0^* \cup B = B$$

8. If $A, 0^* \subset C$, then $A \cup 0^* = 0^* \cup A = A$.

PROOF:

Let $a \in A \cup 0^*$; then $a = b + c$ for some $b \in A$ and $c \in 0^*$ (i.e. $c < 0$), which together with $b + c < b$ implies that $b + c \in A$, i.e. $a \in A$.

$$\text{Hence} \quad A \cup 0^* \subset A \quad (1)$$

Conversely, if $a \in A$, then there exists $d \in A$, where $d \in R$ and $d > a$, such that $c = a - d$, which implies $c < 0$, $c \in 0^*$, and $a = c + d$, so that $a \in A \cup 0^*$.

$$\text{Hence} \quad A \subset A \cup 0^* \quad (2)$$

Hence, by (1) and (2), $A = A \cup 0^*$, where $A \cup 0^* = 0^* \cup A$ by Prob. 4; thus $A \cup 0^* = 0^* \cup A = A$.

9. Prove Th. 5.1.2.5.

PROOF:

If $B \subset A$, then every $b \in B$ implies $b \in A$ and there exists $a \in A$ such that $a \notin B$, i.e. $a \in B'$. Moreover, if there exist $c, d \in A$ such that $a < c < d$, then $-c \in -B$ ($\because a \in B'$). Hence $d - c$ is a positive rational number in the cut $A - B$, i.e. $(A - B) \subset C$ and $A > B$.

Conversely, if $A > B$, i.e. $(a + b) \in (A - B)$ where $a \in A$ and $b < -b'$ for some $b' \in B'$, then $a > -b > b'$ ($\because a + b > 0$ and $b < -b'$) such that $a \notin B$, which together with $a \in A$ implies $A \neq B$. Furthermore, $c < b'$ for every $c \in B$; hence $c < a$, which together with $c \in A$ implies $B \subset A$, completing the proof.

10. Transitivity holds in \mathcal{C} .**PROOF:**

Let $A, B, C \subset \mathcal{C}$, where $A < B$ and $B < C$; then, by Prob. 9, $A \neq B$, $B \neq C$, and $A \subset B$, $B \subset C$, from which it follows that $A \subset C$ (cf. Th. 2.1.5) and that $A \neq C$ (cf. Df. 2.1.8).

Hence $A < C$, which completes the proof.

Second Proof. Since $A < B$, there exists $a \in R$ such that $a \in B$, but $a \notin A$. Likewise $B < C$ implies $b \in R$ such that $b \in C$, but $b \notin B$. Since $a \in B$ and $b \notin B$ imply $a < b$, and this, together with $a \notin A$, in turn implies $b \notin A$, it immediately follows that $b \in C$, but $b \notin A$.

Hence $A < C$.

Note. The relation in \mathcal{C} , however, is neither reflexive nor symmetric, as can be readily verified.

11. Prove Th. 5.1.2.6.

PROOF:

Let $A \neq B$ and also $A \not\leq B$; then there exists some $a \in A$ such that $a \in B'$. This implies $b < a$ for any $b \in B$, which in turn, by Df. 5.1.2.1, implies that $B \subset A$.

Likewise, $A \neq B$ and $A \not\geq B$ dually imply $A \subset B$, from which it follows, together with the first result, that $A \leq B$ and $A \geq B$ imply $A = B$, completing the proof.

12. If $A, B \subset \mathcal{C}$, then $A \cap B \subset \mathcal{C}$.**PROOF:**

Let $A > 0^*$ and $B > 0^*$, which imply $A \neq \emptyset$ and $B \neq \emptyset$; hence $A \cap B \neq \emptyset$. Also $A \cap B \neq R$, since there exist some $a' \notin A$ and $b' \notin B$ such that $a < a'$ and $b < b'$ for every $a \in A$, and $b \in B$, which implies that there exists a positive rational number $a'b' \notin (A \cap B)$ such that $ab < a'b'$ for every $ab \in (A \cap B)$.

Furthermore, (i) if $c \in R$ such that $0 < c < ab$, where $ab \in (A \cap B)$ as above, then $c \in (A \cap B)$, since $c/a < b$ implies $c/a \in B$, and since $c = a(c/a)$.

Also, (ii) there exists $a_1 \in A$ such that $a_1 > a$, since A is a cut. (Or, for the same reason, there exists $b_1 \in B$ such that $b_1 > b$.)

Hence $a_1 b \in (A \cap B)$ (or $ab_1 \in (A \cap B)$) such that $a_1 b > ab$ (or $ab_1 > ab$).

Df. 5.1.2.1 is thus completely satisfied with respect to $A \cap B$; hence $A \cap B \subset \mathcal{C}$.

It is proved likewise that $A \cap B \subset \mathcal{C}$ for (i) $A > 0^*$ and $B < 0^*$ or $A < 0^*$ and $B > 0^*$, and (ii) $A = 0^*$ or $B = 0^*$.

13. Multiplication in \mathcal{C} is both associative and commutative.**PROOF:**

For every $a \in A$, $b \in B$, $c \in C$, where $a, b, c > 0$ and $A, B, C \subset \mathcal{C}$,

(i) $A \cap (B \cap C) = A \cap \{bc\} = \{a(bc)\} = \{(ab)c\} = \{ab\} \cap C = (A \cap B) \cap C$, and

(ii) $A \cap B = \{ab\} = \{ba\} = B \cap A$.

It is proved likewise if (i) $a > 0$, $b > 0$, $c < 0$, (ii) $a > 0$, $b < 0$, $c > 0$, (iii) $a > 0$, $b < 0$, $c < 0$, (iv) $a < 0$, $b > 0$, $c > 0$, (v) $a < 0$, $b > 0$, $c < 0$, (vi) $a < 0$, $b < 0$, $c > 0$, (vii) $a < 0$, $b < 0$, $c < 0$, and (viii) $a = 0$ or $b = 0$ or $c = 0$.

14. Distributive laws hold in \mathcal{C} .**PROOF:**

For every $a \in A$, $b \in B$, $c \in C$, where $A, B, C \subset \mathcal{C}$ and $a, b, c > 0$,

$$a(b + c) = ab + ac$$

which implies that every positive element of $A \cap (B \cup C)$ is a positive element of $(A \cap B) \cup (A \cap C)$. Hence

$$A \cap (B \cup C) \subseteq (A \cap B) \cup (A \cap C) \quad (1)$$

Conversely, if $r \in (A \cap B) \cup (A \cap C)$, it follows from above that $r = s + t$, where $s \in A \cap B$ and $t \in A \cap C$, which implies $s = a_1 b$ and $t = a_2 c$, where $a, a_1, a_2 \in A$, say $a_1 \leq a_2 \leq a$. Hence

$$r = a_1 b + a_2 c \leq ab + ac = a(b + c)$$

which in turn implies $r \in A \cap (B \cup C)$, since $a(b + c) \in A \cap (B \cup C)$. Every positive element of $(A \cap B) \cup (A \cap C)$ is thus also an element of $A \cap (B \cup C)$, i.e.,

$$(A \cap B) \cup (A \cap C) \subseteq A \cap (B \cup C) \quad (2)$$

From (1) and (2) it follows that $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

As in Prob. 13, other cases (e.g. $A > 0^*$, $B > 0^*$, $C < 0^*$, etc.) can be treated likewise, arriving at the same conclusion.

Similarly, $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$, which completes the proof.

15. If $1^* = S_1$ is a D -cut at 1, then $A \cap S_1 = A$ for every $A \subset C$.

PROOF:

Since the case of $A = S_0$ is trivial, let $A > S_0$; then $A \cap S_1 = \{ar\}$, by Df. 5.1.2.3, (ii), where $a \in A$ and $0 < r \in R < 1$. Hence $ar < a$, which implies $ar \in A$, i.e.

$$A \cap S_1 \subseteq A \quad (1)$$

Conversely, since A is a D -cut, $a \in A$ implies some $b \in A$ such that $a < b$, which implies $a \in A \cap S_1$, and also $b \in A \cap S_1$, since $b = b(a/b)$, where $0 < a/b < 1$. Hence

$$A \subseteq A \cap S_1 \quad (2)$$

It follows, then, from (1) and (2), that $A \cap S_1 = A$.

Note. $A \cap S_1 = S_1 \cap A = A$, since multiplication in C is commutative.

16. If A^{-1} is defined to be a set which consists of every element a of R_x , i.e. a rational cut at x , such that $a < 1/b$ for some $b \in A'$, then A^{-1} is a D -cut.

PROOF:

Since $c \in A$ implies $c < b$ for every $b \in A'$, it follows that $1/c > 1/b$ for every $b \in A'$; hence $c \in A$ implies $1/c \notin A^{-1}$, which yields $A^{-1} \neq R$. Also, by hypothesis, $A^{-1} \neq \emptyset$.

Furthermore, (i) $a \in A^{-1}$ and $d < a$, where $d \in R$, do imply $d \in A^{-1}$, and (ii) there exists no largest element in A^{-1} , since $a < 1/b$ for some $b \in A'$ implies, by Th. 5.1.2.2, the existence of a rational number r such that $a < r < 1/b$, which in turn implies $r \in A^{-1}$, satisfying Df. 5.1.2.1 completely.

Hence A^{-1} is a D -cut.

17. Prob. 16 above implies that $A \cap A^{-1} = A^{-1} \cap A = 1^*$.

PROOF:

If $a \in A$, where $A > 0^*$, and $b \in A^{-1}$, then, by Prob. 16, there exists some $a' \in A'$ such that $b < 1/a'$, which implies $0 < ab < a/a' < 1$, i.e. $ab \in 1^*$. Hence

$$A \cap A^{-1} = 1^* \quad (1)$$

Conversely, if $a \in A$, where $A > 0^*$ as before, and $c \in 1^*$ such that $0 < c < 1$, then $1 - c > 0$ and $(1 - c)a > 0$, which by Th. 5.1.2.2 imply some $d \in A$ such that $d + (1 - c)a = a'$ for some $a' \in A'$. Also, evidently, $a < a'$ and $d < a'$, which imply

$$0 < a' - d = (1 - c)a < (1 - c)a'$$

i.e.

$$a' - d + ca' < (1 - c)a' + ca' = a'$$

which implies $ca' < d$, i.e. $a' < d/c$, which in turn implies $c/d < 1/a'$. Hence $c/d \in A^{-1}$, i.e. $c = bd$ for some $b \in A^{-1}$, which implies $c \in A \cap A^{-1}$. But then, by the initial assumption, c is an arbitrary positive element of 1^* ; hence

$$1^* \subset A \cap A^{-1} \quad (2)$$

Hence $A \cap A^{-1} = 1^*$, by (1) and (2) and also, by multiplicative commutativity in C ,

$$A \cap A^{-1} = A^{-1} \cap A = 1^*$$

18. C is a field.

PROOF:

F1 for C is assured by Prob. 2, and **F2, 5** by Prob. 4. Furthermore, **F3** is satisfied by 0^* , and **F4** by Prob. 7.

F6 for C is proved by Prob. 12, **F7, 9** by Prob. 13, and **F8, 10, 11** by Prob. 14, 15, 16, 17 respectively, which exhaust the properties of a field.

Hence C is a field.

19. For any $a, b \in R$ and $S_a, S_b \subset C$, $S_a < S_b$ iff $a < b$.

PROOF:

If $a < b$, then $a \in S_b$, but $a \notin S_a$. Hence $S_a < S_b$.

Conversely, if $S_a < S_b$, then there exists $c \in R$ such that $c \in S_b$, but $c \notin S_a$, which implies $a \leq c < b$, i.e. $a < b$, completing the proof.

20. Prove Th. 5.1.2.7.

PROOF:

Since $A < B$, there exists $a \in R$ such that $a \in B$, but $a \notin A$. Take, then, $c \in R$ such that $a < c$ and $c \in B$. This implies $S_r < B$, since $c \in B$ and $c \notin S_r$. Also $A < S_r$, since $a \in S_r$; but $a \notin A$. Hence, letting $S_r = C$, $A < C < B$.

Note. If A and B are explicitly defined to be rational cuts, this theorem is virtually equivalent to Th. 5.1.1.5. On the other hand, if A and B are merely defined to be contained in C , as is stated in the theorem, they can be irrational as well as rational. Hence the theorem as such is equivalent to the following form: There exists a rational number between any two distinct real numbers (cf. Prob. 26 below).

21. For any cut A , $a \in A$ iff $S_a < A$.

PROOF:

Since $a \notin S_a$ for every $a \in R$, it follows at once that $S_a < A$ if $a \in A$.

Conversely, if $S_a < A$, then there must exist $b \in R$ such that $b \in A$ and $b \notin S_a$. Hence $a \leq b$, which together with $b \in A$ implies $a \in A$, completing the proof.

Note. This theorem has proved that every cut A , representing a real number x , is the set of all rational numbers y such that $y < x$ (cf. Prob. 28 below).

22. If $A \subset C$ and $r \in R^+$, then there exist $p \in A$ and $q \notin A$, but $q \in R$, such that q is not the l.u.b. of A , and that $q - p = r$.

PROOF:

Let $a \in A$ and $a_n = a + nr$, $n = 0, 1, \dots$; then, by Th. 5.1.1.6, there uniquely exists an integer m such that $a_m \in A$ and $a_{m+1} \notin A$. If a_{m+1} as such is not the l.u.b. of A , then the substitution $p = a_m$ and $q = a_{m+1}$ verify the theorem.

If a_{m+1} is the l.u.b. of A , then the substitution $p = a_m + (r/2)$ and $q = a_{m+1} + (r/2)$ verify the theorem.

In either case the theorem is verified, exhausting the cases, and the proof is complete.

23. If A and B are two sets of rational numbers such that every rational number is either in A or in B , but never in both, and that $a < b$ for every $a \in A$ and $b \in B$, then there exists a case where A has the largest and B has not the smallest.

PROOF:

The partition of R defined by hypothesis yields the following four exclusive and exhaustive cases:

- (i) A has the largest number and B has not the smallest.
- (ii) A has not the largest number and B has the smallest.
- (iii) A has not the largest number and B has not the smallest.
- (iv) A has the largest number and B has the smallest.

The first case, then, is to be proved first. Suppose that B contains the smallest number, say b , while $a < 10$ for every $a \in A$; then

$$10 < (10 + b)/2 = r$$

since $b \in B$. Hence $r \notin A$. Also

$$(10 + b)/2 = r < b$$

since $10 < b$. Hence $r \notin B$, contradictory to the initial assumption that every rational number, including r in this context, is either in A or in B .

Hence B must not contain the smallest number if A has the largest.

Note. The case (ii) can be proved likewise, while the case (iii) is proved by Prob. 24 below, and the case (iv) by Prob. 25.

24. Given A and B as defined in Prob. 23 above, show that A has no largest number and B has no smallest.

PROOF:

Let B be the set of all rational numbers b such that $2 < b^2$ and A be the set of all rational numbers a such that $a \notin B$. Then A is evidently a D -cut, since $a < b$ for every $a \in A$.

If, say, $a^2 = 2$, where $a = p/q$ for some $p, q \in I$ such that $(p, q) = 1$, then $p^2 = 2q^2$, which implies that p and q are both even numbers (cf. §2.1, Prob. 4 note), contradictory to the assumption. Likewise $b^2 = 2$ faces a similar contradiction.

Hence neither $a^2 = 2$ nor $b^2 = 2$.

Furthermore, take a rational number c such that $0 < c < 1$ and $c < (2 - a^2)/(2a + 1)$, where a is assumed to be the largest in A such that $a^2 < 2$. Let $b = a + c$; then $a < b$, and

$$b^2 = a^2 + (2a + c)c < a^2 + (2a + 1)c < a + (2 - a) = 2$$

which implies $b \in A$, contradictory to the assumption.

Hence A has no largest number.

Likewise, if b is assumed to be the smallest in B such that $2 < b^2$, take $d \in R$ such that

$$d = b - ((b^2 - 2)/2b) = (b/2) + (1/b)$$

which implies $0 < d < b$ and

$$d^2 = b^2 - (b^2 - 2) + ((b^2 - 2)/2b)^2 > b^2 - (b^2 - 2) = 2$$

which implies $d \in B$, contradictory to the assumption. Hence B has no smallest number.

Thus neither A has the largest number nor B has the smallest.

25. If A and B are defined as in Prob. 23, it is impossible that A has the largest number and B has the smallest.

PROOF:

If A has the largest number, say a_1 , and B has the smallest, say a_2 , and if, by hypothesis, the field is limited to R (or what is the same, the class of rational cuts only), then the density (cf. Th. 5.1.1.5 for R and its equivalent, Th. 5.1.2.7, for rational cuts) of R allows the existence of $a_3 \in R$ such that $a_1 < a_3 < a_2$.

Hence the case (iv) that A has the largest number and B has the smallest is an impossibility in R , thus yielding a leap between A and B .

26. There exists $c \in R$ such that $a < c < b$ for every $a, b \in \bar{R}$.

PROOF:

There are three cases to be considered: (i) both a and b are irrational; (ii) a is rational and b irrational; (iii) a is irrational and b rational.

(i) If a and b are irrational cuts, yielding two pairs of disjoint classes A_1, A_2 and B_1, B_2 respectively (as in Th. 5.1.2.9), then there exist rational numbers in B_1 which are not in A_1 , since $a < b$ by hypothesis. Hence, if c is one of such rational numbers, it follows immediately that $a < c < b$.

(ii) If a is irrational and b rational, then $(a + b)/2$ is also irrational. For, otherwise, it must be rational, say $r = (a + b)/2$, where $r \in R$, which implies $2r - b = a$, contradicting the closure property of rational numbers. Thus, $(a + b)/2$ being irrational, the problem is now very much the same as (i), viz.

$$a < c < (a + b)/2 < b$$

(iii) Likewise, if a is rational and b irrational,

$$a < (a + b)/2 < c < b$$

which completes the proof.

27. There exists an irrational number $d \in \bar{R}$ such that $a < d < b$ for every $a, b \in \bar{R}$.

PROOF:

As in Prob. 26 above, there are three cases to be considered, and also an additional case, viz.

(iv) both a and b are rational.

- (i) If both a and b are irrational, then $(a+b)/2 = c$ may be either rational or irrational. If c is rational, however, $(a+c)/2$ and $(c+b)/2$ must be irrational as in Prob. 26; hence

$$a < (a+b)/2 = c < (a+c)/2 = d < b \quad \text{or} \quad a < (a+b)/2 = c < (c+b)/2 = d' < b$$

If c is irrational outright, then it follows at once

$$a < d \text{ (or } d') < b$$

- (ii) If a is irrational and b rational, then $(a+b)/2 = d$ is also irrational; hence

$$a < d < b$$

- (iii) If a is rational and b irrational, the case is exactly the same as (ii).

- (iv) If both a and b are rational, then take $\sqrt{2}$, which has already been proved to be irrational (cf. Prob. 24 above), so that

$$\sqrt{2}/n < b - a$$

where $n \in \mathbb{N}$ can be made large enough to establish the inequality. Since $a + (\sqrt{2}/2)$ is obviously irrational, it follows at once that

$$a < a + (\sqrt{2}/n) < b$$

completing the proof.

28. Prove Th. 5.1.2.11.

PROOF:

In the context of Prob. 26, let R_1 be the set of rational numbers c such that $c \leq b$ for a given real number b . Then, by this assumption itself, b is an upper bound of R_1 . Also, since it has already been proved by Prob. 26 that no smaller real number (a , in Prob. 26) can be an upper bound of R_1 , b must be the l.u.b. of R_1 , which completes the proof.

Note. The dual (Th. 5.1.2.11a) can be proved likewise, starting with $a \leq c$ in the same context and ending with the conclusion that a is the g.l.b. of R_1 .

29. Prove Th. 5.1.2.13.

PROOF:

- (i) Let \tilde{R}_1 be the set of all real numbers x such that $x < a'$ for some $a' \in A$, and \tilde{R}_2 the set of all other real numbers. Also, let R_1 and R_2 be the sets of all rational numbers in \tilde{R}_1 and \tilde{R}_2 respectively. This implies that the rational number field \mathbb{R} is now divided into two disjoint sets of R_1 and R_2 by a cut at, say, s . Then, by Prob. 26, there exists $r \in R$ such that $a < r < s$ for any $a \in \tilde{R}$ such that $a < s$.

Hence $r \in R_1$, which in turn implies $a \in \tilde{R}_1$.

Likewise, if $b \in \tilde{R}$ such that $s < b$, then, by Prob. 26 again, there exists $s' \in R$ such that $s < r' < b$. Hence $r' \in R_2$, which implies $b \in \tilde{R}_2$.

Hence there does exist a number $s \in \tilde{R}$ such that $a \in \tilde{R}_1$ implies $a < s$ and $b \in \tilde{R}_2$ implies $s < b$.

Conversely, s as such is actually the l.u.b. which satisfies (i). For, if not, by assuming $t \in A$ such that $s < t$, the following inequality

$$s < u = (s+t)/2 < t$$

establishes a contradiction that $u \in \tilde{R}_1$ and $u \in \tilde{R}_2$, both of which follow from the result obtained above.

Furthermore, s is unique. For, if both s_1 and s_2 , say $s_1 < s_2$, are to satisfy (i), s_3 may be assumed by Th. 5.1.2.7 such that

$$s_1 < s_3 < s_2$$

which implies a contradiction that $s_3 \in R_1$ ($\because s_3 < s_2$) and $s_3 \in R_2$ ($\because s_1 < s_3$).

- (ii) Assume, contrary to the conclusion to be drawn, that $c \notin A$ such that $s - p < c$ for any $p > 0$. Then the inequality

$$s - p < v = s - (p/2) < s$$

implies $v \in \tilde{R}_2$, since \tilde{R}_2 by hypothesis is to contain any x such that $x < a'$ for some $a' \in A$, while the same inequality implies $v \in \tilde{R}_1$, since \tilde{R}_1 has already been proved by (i) to contain any number less than s .

Hence the initial assumption is evidently a contradiction, which thus proves (ii) to be valid.

Note. The uniqueness of s may be reaffirmed by (ii). For, if both s and s' , say $s < s'$, are to satisfy (ii), then let $s' - s = p'$, and

$$s < w = s' - (p'/2) < s'$$

introduces a contradiction, since the inequality may be viewed in terms of s and (i), which imply all $a \notin A$ such that $w < a$, while the same may be viewed in terms of s' and (ii), which imply all $a \in A$ such that $w < a$.

It must be also noted that the dual (Th. 5.1.2.13a) can be similarly proved with a very slight modification of Th. 5.1.2.13, as is but natural for a dual.

30. Prove Th. 5.1.2.14.

PROOF:

Prob. 26 has already verified the existence of $r \in R$ such that $a' < r < a < b$ for any $a', a, b \in \bar{R}$. Also, as is quite obvious, there must exist $s \in R$ such that $b < s$. But then, by Th. 5.1.1.6, there exists $n \in I^+$ such that $nr > r$. Hence

$$na > nr > s > b, \quad \text{i.e. } na > b$$

which completes the proof.

31. If $\{a_n\}$ and $\{b_n\}$ are two sequences and if p and q are two positive integers such that $a_{p+n} = b_{q+n}$ for any $n \in N$, then the two sequences either both converge to the same limit or both diverge.

PROOF:

If $\{a_n\}$ converges to a , then every neighborhood of a contains all but a finite number of the terms of $\{a_n\}$, i.e. all but a finite number of the terms of $\{b_n\}$. Hence the alternation of a finite number of terms of a sequence cannot have any effect on the convergence.

The same is true in the case of divergence, since the two sequences either both have the same infinite limit or neither has an infinite limit.

Note. The proof presumed the term “neighborhood” to be intuitively self-evident, but it can be more strictly defined as follows:

If x is a given point and $p > 0$, then the open interval $(x-p, x+p)$, sometimes denoted by $N(x; p)$ or $N(x)$, is called a *neighborhood* of x as center and of radius p .

The neighborhood can be similarly defined in the space of two or three dimensions or more generally of n -dimensions.

32. Any subsequence of a convergent sequence converges, and its limit is the limit of the original sequence.

PROOF:

Let the convergent sequence be $\{a_n\}$, the limit of which is a . Then, since every neighborhood of a contains all but a finite number of the terms of $\{a_n\}$, it must contain all but a finite number of the terms of any subsequence.

Hence any subsequence of $\{a_n\}$ must converge to a , the limit of $\{a_n\}$ itself.

33. The limit of a convergent sequence $\{a_n\}$ is unique.

PROOF:

Suppose $\lim a_n = a$ and $\lim a_n = b$, where $a \neq b$. If, however, the neighborhoods of a and b are made so small that they have no points in common, then each must contain all but a finite number of the terms of $\{a_n\}$, which is evidently a contradiction. Hence $a = b$, which proves the uniqueness of the limit.

Second Proof. Since $a \neq b$, $|a - b| > 0$, and also $|a - b|/2 > 0$. If $\lim a_n = b$, then there must exist $p, q \in N$ such that $|a_n - a| < |a - b|/2$ and $|a_n - b| < |a - b|/2$ for all $n > p$ and $n > q$. Hence

$$|a - b| = |(a - a_n) + (a_n - b)| \leq |a - a_n| + |a_n - b| < |a - b|/2 + |a - b|/2 = |a - b|$$

i.e. $|a - b| < |a - b|$, which is inconsistent. Hence the limit must be unique.

34. Any convergent sequence is bounded.

PROOF:

Let the given sequence be $\{a_n\}$, where $\lim a_n = a$, and take a specific neighborhood of a , say the open interval $(a-1, a+1)$. Then, since this neighborhood contains all but a finite number of the terms of $\{a_n\}$, a suitable enlargement of the neighborhood will contain the remaining terms.

Hence all terms of the sequence is now contained in some finite interval, which makes the sequence bounded, by Df. 5.1.2.19.

35. Given two sequences $A = \{a_n\}$ and $B = \{b_n\}$, the convergence of A and $\lim (a_n - b_n) = 0$ imply the convergence of B and also

$$\lim a_n = \lim b_n$$

and conversely.

PROOF:

Let $\lim a_n = a$ and $p > 0$. Then, by hypothesis, there exist $p, q \in N$ such that $|a_n - a| < p/2$ for all $n > p$ and $|a_n - b_n| < p/2$ for all $n > q$. Hence

$$|b_n - a| = |(b_n - a_n) + (a_n - a)| \leq |b_n - a_n| + |a_n - a| < p/2 + p/2 = p$$

which implies, by Df. 5.1.2.18,

$$\lim a_n = \lim b_n$$

The converse is logically equivalent to Prob. 36 below.

36. The sum (or difference) of two convergent sequences is a convergent sequence, and the limit of the sum is the sum of the limits.

PROOF:

Let the two sequences be $\{a_n\}$ and $\{b_n\}$, where $\lim a_n = a$ and $\lim b_n = b$, and let $p > 0$. If $m \in N$ is chosen to be so large that the following inequalities can be simultaneously held for any $n > m$:

$$|a_n - a| < p/2 \quad \text{and} \quad |b_n - b| < p/2$$

then, by the triangle inequality [cf. Th. 4.1.2.2.13, (i)],

$$|(a_n \pm b_n) - (a \pm b)| = |(a_n - a) \pm (b_n - b)| \leq |a_n - a| + |b_n - b| < p/2 + p/2 = p$$

Hence, by Df. 5.1.2.18,

$$\lim (a_n \pm b_n) = \lim a_n \pm \lim b_n$$

Note. The result can be readily generalized to the sum (or difference) of any finite number of sequences (cf. Supplementary Problem 5.4).

37. If $\{a_n\}$ is a convergent sequence, and if $\{b_n\}$ converges to 0, then $\{a_n b_n\}$ converges to 0.

PROOF:

Since $\{a_n\}$ is bounded, there exists $k \in I^+$ such that $|a_n| \leq k$ for all $n \in N$. Let $p > 0$, and take $m \in N$ so large that $|b_n| < p/k$ for $n > m$. Then, for $n > m$,

$$|a_n b_n - 0| = |a_n b_n| = |a_n| \cdot |b_n| < (p/k) \cdot k = p$$

Hence, by Df. 5.1.2.18, $\lim a_n b_n = 0$, which completes the proof.

38. The product of two convergent sequences is again a convergent sequence, and the limit of the product is the product of the limits of the two sequences.

PROOF:

Assume as in Prob. 36; then

$$a_n b_n - ab = (a_n - a)b_n + (b_n - b)a$$

where, by Prob. 37, both sequences $(a_n - a)b_n$ and $(b_n - b)a$ converge to 0 ($\because \lim (a_n - a) = 0$ and $\lim (b_n - b) = 0$ by the initial assumption).

Hence, by Prob. 36-37,

$$\lim (a_n b_n) = ab = (\lim a_n) \cdot (\lim b_n)$$

Note. Like Prob. 36, Prob. 38 can be readily generalized to the product of any finite number of sequences. Furthermore, the proof itself can be carried out independently of Prob. 37, viz.:

Second Proof. Since $\lim a_n = a$ by hypothesis, there exists $p \in N$ such that $|a_n - a| < 1$ for all $n > p$, which implies

$$|a_n| = |(a_n - a) + a| \leq |a_n - a| + |a| < 1 + |a|$$

Let a' be the largest of $|a_1|, |a_2|, \dots, |a_p|, 1 + |a|$, and also $a' + 1 = c$; then $c > 1 > 0$, which implies $|a_n| < c$ for all n . Hence the sequence $\{a_n\}$ is bounded.

Furthermore, choose d suitably, so that $d > |b|$, say $d = |b| + 1$, which implies $d > 0$. Since $\lim a_n = a$ and $\lim b_n = b$ by hypothesis, there exist $s, t \in N$ for $p > 0$ such that $|a_n - a| < p/2d$ and $|b_n - b| < p/2d$ for all $n > s$ and $n > t$. Hence

$$\begin{aligned} |a_n b_n - ab| &= |(a_n b_n - a_n b) + (a_n b - ab)| \leq |a_n b_n - a_n b| + |a_n b - ab| \\ &= |a_n| |b_n - b| + |b| |a_n - a| < d(p/2d) + d(p/2d) = p \end{aligned}$$

Hence, by Df. 5.1.2.18,

$$\lim (a_n b_n) = ab = (\lim a_n)(\lim b_n)$$

39. The quotient of two convergent sequences, where the denominators and their limits are not zero, is a convergent sequence, and the limit of the quotient is the quotient of the limits of the sequences.

PROOF:

Assume as in Prob. 36; then

$$a_n/b_n = (a_n) \cdot (1/b_n)$$

and the proof is complete, by Prob. 38, if it is proved that

$$\lim 1/b_n = 1/b, \quad \text{i.e.} \quad \lim ((1/b_n) - (1/b)) = \lim ((b - b_n)/b) \cdot (1/b_n) = 0$$

Since the sequence $\{(b - b_n)/b\}$ converges to 0, by Prob. 37 and the initial assumption, the proof is then complete if the sequence $\{1/b_n\}$ is proved to be bounded. But, since $b \neq 0$ by hypothesis, two neighborhoods of 0 and b can be chosen such that they have no points in common, and since $\{b_n\}$ converges to b by hypothesis, the neighborhood of b contains all but a finite number of the terms of $\{b_n\}$, which in turn implies that the neighborhood of 0 may contain only a finite number of these terms.

Furthermore, since $b_n \neq 0$ for all $n \in N$, there must exist a smaller neighborhood of 0 which excludes all terms of $\{b_n\}$. If this neighborhood is the open interval $(-p, p)$, where $p > 0$, then for all $n \in N$,

$$|b_n| \geq p, \quad \text{i.e.} \quad |1/b_n| \leq 1/p$$

which implies that $\{1/b_n\}$ is bounded, completing the proof.

Note. As was possible in Prob. 38, the proof above may be carried out without resorting to Prob. 37, viz.:

Second Proof. There exists $s \in N$ such that $|b_n - b| < |b|/2$ for all $s < n$. For, if not, there exists $t \in N$ such that $t < s$ and $|b_t| \leq |b|/2$, which implies

$$|b| = |(b - b_t) + b_t| \leq |b - b_t| + |b_t| < |b|/2 + |b|/2 = |b|$$

i.e. $|b| < |b|$, which is absurd.

Hence there must exist $s \in N$ such that $|b_n| > |b|/2$ for all $s < n$.

Furthermore, by Df. 5.1.2.19, the sequence $\{a_n\}$ is bounded as it is convergent. Hence there exists $r \in \mathbb{R}^+$ such that $|a_n| < r$ for all $n \in N$.

Finally, since $\lim a_n = a$ and $\lim b_n = b$, there exist $t, t' \in N$ such that $|a_n - a| < p|b|/2$ for all $t < n$ and $|b_n - b| < pb^2/4r$ for all $t' < n$, which together imply

$$\begin{aligned} |a_n/b_n - a/b| &= |(a_n b - ab_n)/b_n b| = |(a_n b - a_n b_n) + (a_n b_n - ab_n)| / |b_n b| \\ &\leq |a_n b - a_n b_n| / |b_n b| + |a_n b_n - ab_n| / |b_n b| = |a_n| |b - b_n| / |b_n b| + |a_n - a| / |b| \\ &< r(r b^2/4r) / ((|b|/2) |b|) + (p|b|/2) / |b| = p \end{aligned}$$

which in turn implies, by Df. 5.1.2.18,

$$\lim (a_n/b_n) = a/b = (\lim a_n)/(\lim b_n)$$

40. If $\lim a_n > \lim b_n$, then there exists $m \in N$ such that $a_n - b_n > p$, where $p > 0$, for all $n > m$, and conversely.

PROOF:

Let $a > b$ and $p = (a - b)/3 > 0$; then, by hypothesis, there exist $s, t \in N$ such that $|a_n - a| < p$ and $|b_n - b| < p$ for all $n > s$ and $n > t$. If $a_n - b_n \leq p$, then

$$a - b = (a - a_n) + (a_n - b_n) + (b_n - b) < p + p + p = 3p = a - b$$

i.e. $a - b < a - b$, which is absurd.

Hence $a_n - b_n > p$ for all $n < n'$, where n' is the smaller of s and t .

Conversely, if $a_n - b_n \geq 0$ for all $n > n'$, then $a < b$ implies the existence of $p > 0$ and $s \in N$ such that $b_n - a_n > p > 0$ for all $n > s$. Let $n > n'$ and $n > s$; then $a_n \geq b_n$ and yet $b_n < a_n$, which is again absurd.

Hence it must be the case that $a \geq b$, which completes the proof.

41. Prove Th. 5.1.2.22.

PROOF:

Let the given sequence be $\{a_n\}$ and $\lim a_n = a$. Then there exists $m \in N$ for any $p \in F$, where $p > 0$, such that $|a_n - a| < p/2$ for all $n > m$. If $s > m$ and $t > m$, then by the triangle inequality (cf. Prob. 36 above),

$$|a_m - a_n| = |(a_m - a) - (a_n - a)| \leq |a_m - a| + |a_n - a| < p/2 + p/2 = p$$

Hence, by Df. 5.1.2.21, $\{a_n\}$ is a C -sequence.

Note. This theorem may hold even for some divergent sequences if the Axiom of Completeness (cf. Df. 5.1.2.15) is to be ignored. E.g. a sequence of rational numbers converging to $\sqrt{2}$ (an irrational number) does satisfy this theorem, but does not converge within the field of rational numbers itself.

42. The concept of limits and C -sequences in an ordered field \bar{F} can be applied to a subfield \bar{F}' of \bar{F} iff \bar{F} is Archimedean-ordered.

PROOF:

Assume that \bar{F} is not Archimedean-ordered; then, by Prob. 30, there must exist an element $a \in \bar{F}$ such that $n \leq a$ for some $n \in N$, which implies $r < n \leq a$, i.e. $r < a$, for every $r \in R$, since the rational number field R is Archimedean-ordered (cf. Th. 5.1.1.6). If a is multiplied by $1/ar > 0$, then $1/a < 1/r$, i.e. $0 < 1/a < b$, where $1/r = b \in R^+$. Now, let $1/a = p > 0$; then $|1/c - 1/d| > p$ for any $c, d \in \bar{F}$ and $c \neq d$. Hence there does not exist $n' \in N$ such that $|1/c - 1/d| < p$ for every $c, d > n'$ if \bar{F} is not Archimedean-ordered. That is, say, a rational sequence $\{1/n\}$, $n = 1, 2, \dots$, which evidently converges to zero, is a C -sequence in R (which is of course a subfield of \bar{F}), and yet does not converge in \bar{F} , which is absurd.

Hence \bar{F} must be Archimedean-ordered.

Conversely, if \bar{F} is Archimedean-ordered, then the convergence of a C -sequence, $\lim a_n = a$, is assured in \bar{F}' if the same holds in \bar{F} . Since \bar{F} is Archimedean-ordered, there exists $n \in N$ such that $n > 1/p$, i.e. $0 < 1/n = p' < p$, where $p' \in R$, hence $p' \in \bar{F}'$. Then, since $\lim a_n = a$ in \bar{F}' , there must exist $n' \in N$ such that $|a_n - a| < p' < p$ for every $n < n'$.

Hence $\lim a_n = a$ in \bar{F} , which implies $\lim a_n = a$ in \bar{F}' iff the same holds in \bar{F} , completing the proof.

43. An ordered field \bar{F} , which contains the rational field R , is Archimedean-ordered iff each element of \bar{F} is the limit of a rational sequence.

PROOF:

Let $a \in \bar{F}$, where a is the limit of a rational sequence $\{a_n\}$. This implies the existence of $k \in N$ such that $|a_k - a| < 1$ and

$$a \leq |a| = |(a - a_k) + a_k| \leq |a - a_k| + |a_k| < 1 + |a_k|$$

where evidently $1 + |a_k| \in R$, by the initial assumption. But then, since R is an Archimedean-ordered field (cf. Th. 5.1.1.6), there exists $n \in N$ such that $n > 1 + |a_k|$. Hence $n > a$, which implies by Df. 4.1.2.4.12 that \bar{F} is Archimedean-ordered.

Conversely, if \bar{F} is Archimedean-ordered, then there exist $s, t \in N$ for any $n \in N$ and any $a \in \bar{F}$ such that

$$s(1/n) > a \quad \text{and} \quad t(1/n) > -a, \quad \text{i.e. } (-t)(1/n) < a$$

which implies that the set M of integers, in which $r(1/n) \leq a$ for every $r \in M$, is bounded above. Hence M must contain the largest integer m such that

$$m/n \leq a < (m+1)/n$$

i.e. $0 \leq a - (m/n) < 1/n$.

Let $m/n = a_n$; then, since there exists $n' \in N$ such that $n' > 1/p$ for $p \in \bar{F}^+$ (i.e. $p \in \bar{F}$ and $p > 0$) and

$$|a_n - a| = a - a_n < 1/n < 1/n' < 0$$

for every $n > n'$.

This implies $\lim a_n = a$ for every $a \in \bar{F}$, completing the proof.

44. If \bar{R} and \bar{R}' are two real number fields, there exists one, and only one, order-isomorphism M (cf. Th. 4.1.2.3.8) of \bar{R} into \bar{R}' , in which the rational number field R is one and the same.

PROOF:

Since \bar{R} is Archimedean-ordered, it follows from Prob. 43 that there exists a rational sequence, say $\{a_n\}$, such that $\lim a_n = a$ for every $a \in \bar{R}$, which implies that $\{a_n\}$ is a C -sequence in \bar{R} . Hence, by Prob. 42, $\{a_n\}$ is also a C -sequence in the rational number field R , where of course $R \subset \bar{R}$ and $R \subset \bar{R}'$. Since \bar{R}' is also Archimedean ordered, by hypothesis, $\{a_n\}$ is a C -sequence in \bar{R}' , too, and consequently $\lim a_n = a'$ for $a' \in \bar{R}'$, where a' is independent from any specific choice of the rational sequence $\{a_n\}$ for the following reason:

If there exists a rational sequence, say $\{b_n\}$, such that $\lim b_n = a$, then $\lim a_n = \lim b_n$ in \bar{R} and \bar{R} , which implies $\lim (a_n - b_n) = 0$ in \bar{R} and likewise in \bar{R}' , which in turn implies $\lim a_n = \lim b_n = a'$.

If f is the mapping of \bar{R} into \bar{R}' , i.e. $a' = f(a)$, then $\lim a_n = a$, where $a \in \bar{R}$, entails $a_n = a$ for every $n \in N$. Hence $f(a) = a$ or $f(a') = a'$; i.e. R is not affected by the mapping f .

On the other hand, if $\lim a_n = a$ and $\lim b_n = b$ where $a \neq b$ in \bar{R} , then $\lim (a_n - b_n) \neq 0$ in \bar{R}' , which implies $\lim a_n \neq \lim b_n$ in \bar{R}' , i.e. $a' = f(a) \neq f(b) = b'$ or $a' \neq b'$ in \bar{R} . Hence the mapping f of \bar{R} into \bar{R}' is 1-1.

Furthermore, for the same mapping f ,

$$\begin{aligned} f(a+b) &= f(\lim a_n + \lim b_n) = f(\lim (a_n + b_n)) = \lim f(a_n + b_n) \\ &= \lim (f(a_n) + f(b_n)) = \lim f(a_n) + \lim f(b_n) \\ &= f(\lim a_n) + f(\lim b_n) = f(a) + f(b) \end{aligned}$$

and likewise

$$f(a \cdot b) = f(a) \cdot f(b)$$

Hence the mapping f is an isomorphism.

The mapping is also an order-isomorphism. For, if

$$\lim a_n = a < b = \lim b_n$$

in \bar{R} , then there exists $n' \in N$ such that $a_n < b_n$ for every $n > n'$, which implies that $\lim a_n < \lim b_n$ in \bar{R}' , by Prob. 40, i.e.

$$a' = f(a) \leq f(b) = b'$$

But $a \neq b$, as the above implies $f(a) \neq f(b)$. Hence $a < b$ implies $a' < b'$, which completes the proof.

45. The mapping f , defined by Prob. 44, is unique.

PROOF:

Assume g to be also an order-isomorphism of \bar{R} into \bar{R}' . Then g maps the subfield R of \bar{R} into a rational field R' contained in \bar{R}' , which cannot be affected by the mapping, as was proved by Prob. 44. Hence $R' = R$, i.e. $g(r) = r$ for every $r \in R$ in \bar{R} .

If $f \neq g$, then there must exist an element $a \in \bar{R}$ such that $a' = f(a) \neq g(a) = b'$. Say $a' < b'$; then let $b' - a' < 1/n$, where $n \in N$. This implies

$$m/n \leq a' < c = (m+1)/n$$

for some $m \in N$, i.e., $a' < c = (m/n) + (1/n) < a' + (b' - a') = b'$

But $c \in R$ and, by Prob. 44, $c = f(c)$, which implies $f(d) = a' < c$, while $c = g(c)$ implies

$$g(d) = b' < c = g(c)$$

i.e. $b' < c$, which is contradictory to $a' < c < b'$.

Hence it must be the case that $f = g$, completing the proof.

Note. This theorem thus assures the uniqueness of the real number field \bar{R} , but the existence of \bar{R} is revealed nowhere within the frame of the theorem. It can be constructed *in concreto*, of course, on the strength of Th. 5.1.2.11 and Th. 5.1.2.22, which Cantor himself used for the construction of \bar{R} .

§5.1.3 Complex Numbers

Df. 5.1.3.1 A complex number z is an ordered pair (or Hamilton's number couple) of the form (x, y) , where $x, y \in \bar{R}$, the real number field, obeying the following binary operative rules: for every $a, b, c, d \in \bar{R}$,

- (i) $(a, b) + (c, d) = (a + c, b + d)$,
- (ii) $(a, b) \cdot (c, d) = (ac - bd, ad + bc)$.

The former, x , of the pair $(x, y) = z$ is sometimes called the real part (or component) of the pair, denoted by $R(z)$ or $\text{Re}(z)$, while the latter, y , is called the imaginary part (or component), denoted by $I(z)$ or $\text{Im}(z)$, since (x, y) may be replaced by its equivalent: $x + iy$, where $i = \sqrt{-1}$.

In this definition the knowledge of i is taken for granted, in particular the following property:

Df. 5.1.3.1a For every $(a, b) \in C$, $(a, b) = a + ib = 0$ iff $a = 0$ and $b = 0$. (Cf. Prob. 1 note.)

Stated otherwise, for every $a \in \bar{R}$ and every $ib \in C$, a and ib are linearly independent (cf. Df. 4.1.3.2.4); viz. $ax + iby = 0$ for every $x, y \in \bar{R}$ iff $x = y = 0$.

It must be noted that Hamilton's quaternions (or quadruples, cf. Df. 4.1.3.1.3) are capable of subsuming complex numbers (i.e. Hamilton number couples), viz.

$$\begin{aligned} (a, b) + (c, d) &= (a, b, 0, 0) + (c, d, 0, 0) = (a + c, b + d, 0, 0) \\ (a, b) \cdot (c, d) &= (a, b, 0, 0) \cdot (c, d, 0, 0) = (ac - bd, ad + bc, 0, 0) \end{aligned}$$

While quaternions form only a sfield, however, complex numbers form a field, viz.:

Th. 5.1.3.2 A set C , each element of which is of the form defined by Df. 5.1.3.1, forms a field. (Cf. Prob. 1.)

Df. 5.1.3.3 The field C , established by Th. 5.1.3.2 is called the *complex number field*.

The field C is evidently embedded in the sfield \bar{Q} of quaternions, as has already been pointed out above; so is the real number field \bar{R} embedded in C (cf. Prob. 2-3), with the following clear-cut provision:

Th. 5.1.3.4 The complex number field C is not ordered. (Cf. Prob. 7.)

Here can be drawn a line of demarcation between \bar{R} and C , to which every operative property but the order of \bar{R} can be carried over. It must be remembered, too, that C is isomorphic to the quotient ring $I\{x\}/(x^2+1)$ (cf. §4.2.3, Prob. 5, and also Prob. 5 below). As will be seen later (cf. Th. 5.2.1.19), C is actually isomorphic to $\bar{R}\{x\}/(x^2+1)$.

Also, as the student has already been introduced through College Algebra and Trigonometry, C is isomorphic to the so-called *complex-plane* (or the Gauss-Argand plane, or to be more historically exact, the Wessel-Argand-Gauss plane), where the orthogonal axes X (the *real axis*) and Y (the *imaginary axis*) through the origin O (cf. Fig. 5.1.3.a) correspond to the real and imaginary parts of complex numbers; viz. a point P in the plane 1-1 corresponds to a complex number: $z = x + iy$.

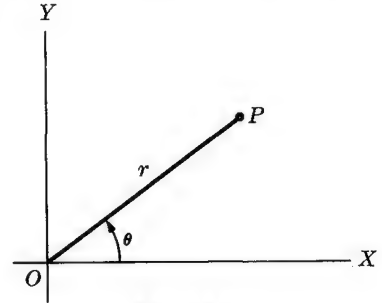


Fig. 5.1.3.a

Df. 5.1.3.5 A complex number $z = x + iy$ is uniquely determined by the *polar coordinates* r and θ , where r is the distance between O and P (cf. Fig. 5.1.3.a) and θ the angle from the real axis to the segment OP , such that

$$r \equiv |z| = \sqrt{x^2 + y^2}, \quad \theta \equiv \arg z = \tan^{-1}(y/x)$$

The polar coordinates r and θ in this context are called the *absolute value* (or *modulus*) and the *argument* (or *amplitude*), respectively, of the complex number z .

It immediately follows that $x = r \cos \theta$ and $y = r \sin \theta$, hence $z = r \cos \theta + ir \sin \theta$ (or more concisely, $z = r \operatorname{cis} \theta$), and that there exists an isomorphism between (r, θ) and (x, y) under rational operations, as can be readily verified through the familiar *de Moivre's theorem*, viz.:

Th. 5.1.3.6 (by de Moivre). If $n \in I$, then

$$(r \operatorname{cis} \theta)^n = r^n \operatorname{cis} (n\theta)$$

(Cf. Prob. 9.)

The triangle inequality (cf. Th. 4.1.2.2.13) as well as the Schwarz inequality (cf. Prob. 19 below) also holds in C with a slight modification, i.e. in terms of moduli:

Th. 5.1.3.7 If $z_1, z_2 \in C$, then

$$|z_1| - |z_2| \leq |z_1 - z_2| \leq |z_1| + |z_2|$$

(Cf. Prob. 12.)

From the theory of equations (cf. Prob. 21-30 below) and others, two complex numbers differing only in the signs of their imaginary parts are specifically defined as follows:

Df. 5.1.3.8 For every $x + iy = z \in C$, the *conjugate* of z , denoted by \bar{z} , is $x - iy = \bar{z} \in C$.

Directly from this definition there follow (i) $|z| = |\bar{z}|$, (ii) $z\bar{z} = |z|^2$, which is sometimes called the *norm* of z , (iii) $\bar{\bar{z}} = z$, (iv) $\overline{z_1 + z_2} = \bar{z}_1 + \bar{z}_2$, (v) $\overline{z_1 \cdot z_2} = \bar{z}_1 \cdot \bar{z}_2$, (cf. Prob. 13-14), which also establish the following theorem:

Th. 5.1.3.9 The mapping $V: z \leftrightarrow \bar{z}$, where $z, \bar{z} \in C$, is an automorphism. (Cf. Prob. 32.)

Note, also, that each real number maps into itself through V , which can be repeated, yielding the following result.

Th. 5.1.3.10 The double mapping V^2 is an identity transformation (cf. Df. 2.2.2.16). (Cf. Prob. 33.)

Note. An isomorphism is sometimes called an *involution* (cf. §2.4.2, Prob. 5) if, when applied twice, it has the effect of an identity transformation; V , then, is an involution in the above context, where, geometrically, it is merely a reflection of z in the X -axis, as is quite obvious in the location of \bar{z} with respect to z .

Of some elementary mappings of the complex plane into itself, the following transformation has particularly interesting properties.

Df. 5.1.3.11 The mapping f defined by the linear fraction of the form

$$W = f(z) = (az + b)/(cz + d)$$

where $ad - bc \neq 0$ and $a, b, c, d \in C$, is called the *Möbius* (or *linear* or *homographic*) transformation. (Cf. Prob. 34-39.)

Th. 5.1.3.12 The set M of all Möbius mappings forms a transformation group. (Cf. Prob. 40.)

Note. M represents some familiar transformations, of infinite varieties, e.g. (i) translations when $w = z + b$ (i.e. $a = d = 1$, $c = 0$); (ii) rotations when $w = az$, $a = \text{cis } \varphi$, φ being the angle of rotation; (iii) dilations when $w = az$, $a > 0$ and $a \neq 1$; (iv) inversions when $w = 1/z$, etc.

Solved Problems

1. $a + ib = c + id$, where $a, b, c, d \in \bar{R}$ and $i = \sqrt{-1}$, iff $a = c$ and $b = d$.

PROOF:

If $a = c$ and $b = d$, then simply by substitution (cf. MTh. 1.1.1.9) $a + ib = c + id$.

Conversely, if $a + ib = c + id$ and also $b = d$, then $ib = id$, which immediately implies, by hypothesis, $a = c$. If, however, $a + ib = c + id$ while $b \neq d$, then $i = (a - c)/(d - b)$, which implies $i \in \bar{R}$, since $a, b, c, d \in \bar{R}$. But $i \notin \bar{R}$, by the definition of i itself.

Hence, in either case, $a + ib = c + id$ implies $a = c$ and $b = d$.

Note. If the concept of "linear independence" (cf. Df. 4.1.3.2.4 and also Prob. 3 below) is presumed in this context, it follows at once that $(a - c) + i(b - d) = 0$, which implies $a = c$ and $b = d$, and conversely.

2. Prove Th. 5.1.3.2.

PROOF:

The set C does satisfy all the properties of a field (cf. Th. 4.1.2.4.1), viz. for every $(a, b), (c, d), (e, f) \in C$:

C1, 6. By Df. 5.1.3.1 itself.

C2. $(a, b) + ((c, d) + (e, f)) = (a, b) + (c + e, d + f) = (a + c + e, b + d + f) = ((a + c) + e, (b + d) + f)$
 $= (a + c, b + d) + (e, f) = ((a, b) + (c, d)) + (e, f)$

C3. $(0, 0)$. ($\because (a, b) + (0, 0) = (a + 0, b + 0) = (a, b)$.)

C4. $-(a, b)$ for every $(a, b) \in C$

($\because (a, b) + (-(a, b)) = (a, b) + (-a, -b) = (a - a, b - b) = (0, 0) = (-(a, b)) + (a, b)$.)

$$\text{C5. } (a,b) + (c,d) = (a+c, b+d) = (c+a, d+b) = (c,d) + (a,b).$$

$$\begin{aligned} \text{C7. } (a,b)((c,d)(e,f)) &= (a,b)(ce-df, cf+de) = (ace-adf-bcf-bde, acf+ade+bce-bdf) \\ &= (ace-bde-adf-bcf, acf-bdf+ade+bce) = (ac-bd, ad+bc)(e,f) \\ &= ((a,b)(c,d))(e,f) \end{aligned}$$

$$\begin{aligned} \text{C8. } (a,b)((c,d) + (e,f)) &= (a,b)(c+e, d+f) = (ac+ae-bd-bf, ad+af+bc+be) \\ &= (ac-bd+ae-bf, ad+bc+af+be) \\ &= (ac-bd, ad+bc) + (ae-bf, af+be) = (a,b)(c,d) + (a,b)(e,f) \end{aligned}$$

Likewise $((a,b) + (c,d))(e,f) = (a,b)(e,f) + (c,d)(e,f)$.

$$\text{C9. } (1,0). \quad (\because (a,b)(1,0) = (a \cdot 1 - b \cdot 0, a \cdot 0 + b \cdot 1) = (a,b).)$$

$$\text{C10. } (a,b)^{-1} = (a/(a^2+b^2), -b/(a^2+b^2)). \quad (\text{Cf. Prob. 3 below.})$$

$$\text{C11. } (a,b)(c,d) = (ac-bd, ad+bc) = (ca-db, cb+da) = (c,d)(a,b)$$

3. Given $(a,b)(x,y) = (c,d)$, where $(a,b), (c,d), (x,y) \in C$ and not simultaneously $a=0, b=0$, find (x,y) .

Solution:

By Df. 5.1.3.1 and hypothesis,

$$(a,b)(x,y) = (ax-by, ay+bx) = (c,d) \quad \text{or} \quad (ax-by) + i(ay+bx) = c + id$$

where, by definition, $a, b, c, d, x, y \in \bar{R}$, which implies $(ax-by), (ay+bx) \in R$. Then, since any real number and $i = \sqrt{-1}$ are linearly independent (cf. Df. 4.1.3.2.4), i.e. $px+iq \neq 0$ for any nonzero $p, q, x \in \bar{R}$, it follows that

$$ax-by = c \quad \text{and} \quad ay+bx = d$$

Hence, solving these equations simultaneously for x and y ,

$$x = (ac+bd)/(a^2+b^2), \quad y = (ad-bc)/(a^2+b^2)$$

Note. A direct verification yields

$$\begin{aligned} (a,b)((ac+bd)/(a^2+b^2), (ad-bc)/(a^2+b^2)) \\ = ((a^2c+abd-bad+b^2c)/(a^2+b^2), (a^2d-abc+bac+b^2d)/(a^2+b^2)) = (c,d) \end{aligned}$$

and if $(c,d) = (1,0)$, then evidently

$$(x,y) = (a/(a^2+b^2), -b/(a^2+b^2))$$

verifying C9-10 in Prob. 2 above.

4. The real number field \bar{R} is embedded in the complex number field C .

PROOF:

Let $a \leftrightarrow (a,0)$ and $b \leftrightarrow (b,0)$, where $a, b \in \bar{R}$ and $(a,0), (b,0) \in C'$; then

$$(a,0) + (b,0) = (a+b, 0) \leftrightarrow a+b \quad \text{and} \quad (a,0) \cdot (b,0) = (a \cdot b, 0) \leftrightarrow a \cdot b$$

which proves an isomorphism of \bar{R} into a subfield C' of C , where every element of C' is uniquely of the form $(x,0)$ for every $x \in \bar{R}$.

Hence \bar{R} is now embedded in C .

5. The real number field \bar{R} and $i = \sqrt{-1}$ form a minimal field containing \bar{R} , i.e. the complex number field C , iff every element of C is uniquely of the form $x+iy$, where $x, y \in \bar{R}$.

PROOF:

Assume that C contains a subfield B which consists of \bar{R} and $j = \sqrt{-1}$. Then, since $i^2 = j^2 = -1$, it follows that

$$(i+j)(i-j) = i^2 - ij + ji - j^2 = 0$$

which implies $i+j=0$ or $i-j=0$, i.e. $i = \pm j$, since, by Df. 5.1.2.4.1, C is not to contain any zero-divisors. Hence $z \in B$, where $z = x+iy = x \pm jy$, if $z \in C$, and conversely, which evidently implies $B=C$, proving that C is a minimal field.

Conversely, if C is a minimal field, then every element of C is uniquely of the given form $x+iy$. For, if $a, b \in C$, where $a = x+iy$ and $b = x'+iy'$, then, by Prob. 1 above and Df. 5.1.3.1,

- (i) $x+iy = x'+iy'$ iff $x=x'$ and $y=y'$;
- (ii) $(x+iy) + (x'+iy') = (x+x') + i(y+y')$;
- (iii) $(x+iy) \cdot (x'+iy') = (xx'-yy') + i(xy'+x'y)$.

This completes the proof.

6. The complex number field C , except its possible isomorphisms, is unique.

PROOF:

Let C and C' be two complex number fields, where $i = i' = \sqrt{-1}$ if $i \in C$ and $i' \in C'$. Then, by Prob. 5, every element of C and C' is to be uniquely of the form $x + iy$ and $x + i'y$ respectively, where $x, y \in \bar{R}$, which thus implies the 1-1 mapping f of C into C' .

Furthermore, if $z_1, z_2 \in C$, where $z_1 = a + ib$ and $z_2 = c + id$, then $f(z_1) = a + i'b$, $f(z_2) = c + i'd$, and

$$\begin{aligned} f(z_1 + z_2) &= f((a + ib) + (c + id)) = f((a + c) + i(b + d)) \\ &= (a + c) + i'(b + d) = (a + i'b) + (c + i'd) = f(z_1) + f(z_2) \end{aligned}$$

Likewise $f(z_1 \cdot z_2) = f(z_1) \cdot f(z_2)$.

Hence f is an isomorphism, which maps i into i' and a into a itself for every $a \in \bar{R}$. This completes the proof.

7. Prove Th. 5.1.3.4.

PROOF:

Assume that C is ordered. Then, by Th. 4.1.2.2.9, $i \neq 0$ implies $i^2 > 0$, contradictory to the definition: $i^2 = -1 < 0$.

Hence C cannot be an ordered field.

8. Express the following complex numbers in their respective polar forms:

$$(i) -1 + i, \quad (ii) -2 - 3i, \quad (iii) i/\sqrt{3} - i$$

Solution:

- (i) Since, by Df. 5.1.3.5, $z = x + iy = -1 + i$, i.e. $x = -1$ and $y = 1$, in this context, it follows at once that $r = |z| = \sqrt{(-1)^2 + 1^2} = \sqrt{2}$. Furthermore, $\theta = \arg z = \tan^{-1}(y/x) = \tan^{-1}(-1)$, i.e. $\theta = 3\pi/4 + 2k\pi$ or, taking its principal value, $\theta = 3\pi/4$, since $\cos \theta = -1/\sqrt{2}$ and $\sin \theta = 1/\sqrt{2}$.

Hence $z = \sqrt{2} \operatorname{cis}(3\pi/4)$.

- (ii) Likewise, $r = \sqrt{13}$, which implies $\cos \theta = -2/\sqrt{13}$ and $\sin \theta = -3/\sqrt{13}$, which in turn implies $\theta = 2k\pi + (\pi + \tan^{-1}(3/2))$ or, taking its principal value, $\theta = \pi + \tan^{-1}(3/2)$.

Hence $z = \sqrt{13} \operatorname{cis}(\pi + \tan^{-1}(3/2))$.

- (iii) Likewise $z = (2/\sqrt{3}) \operatorname{cis}(-\pi/3)$.

9. Prove Th. 5.1.3.6.

PROOF:

Let $z_1 = r_1 \operatorname{cis} \theta_1$ and $z_2 = r_2 \operatorname{cis} \theta_2$; then

$$\begin{aligned} z_1 z_2 &= r_1 \operatorname{cis} \theta_1 \cdot r_2 \operatorname{cis} \theta_2 = r_1 r_2 (\cos \theta_1 + i \sin \theta_1)(\cos \theta_2 + i \sin \theta_2) \\ &= r_1 r_2 ((\cos \theta_1 \cos \theta_2 - \sin \theta_1 \sin \theta_2) + i(\sin \theta_1 \cos \theta_2 + \cos \theta_1 \sin \theta_2)) \\ &= r_1 r_2 (\cos(\theta_1 + \theta_2) + i \sin(\theta_1 + \theta_2)) = r_1 r_2 \operatorname{cis}(\theta_1 + \theta_2) \end{aligned}$$

and in general, by induction, for $z_1 = r_1 \operatorname{cis} \theta_1$, $z_2 = r_2 \operatorname{cis} \theta_2$, ..., $z_n = r_n \operatorname{cis} \theta_n$,

$$z_1 z_2 \cdots z_n = r_1 \operatorname{cis} \theta_1 \cdot r_2 \operatorname{cis} \theta_2 \cdots r_n \operatorname{cis} \theta_n = r_1 r_2 \cdots r_n \operatorname{cis}(\theta_1 + \theta_2 + \cdots + \theta_n)$$

If $z_1 = z_2 = \cdots = z_n = z$, then $r_1 = r_2 = \cdots = r_n = r$ and $\theta_1 = \theta_2 = \cdots = \theta_n = \theta$, which imply

$$z^n = r^n \operatorname{cis}(n\theta)$$

which proves the case for $n \in I^+$.

On the other hand, if $-n \in I^+$, then, since

$$(\operatorname{cis} \theta)^{-1} = 1/(\operatorname{cis} \theta) = (\cos \theta - i \sin \theta)/(\cos \theta + i \sin \theta) = \cos \theta - i \sin \theta = \operatorname{cis}(-\theta)$$

it follows at once that, applying the first result,

$$(\operatorname{cis} \theta)^{-n} = ((\operatorname{cis} \theta)^{-1})^n = (\operatorname{cis}(-\theta))^n = \operatorname{cis}(-n\theta)$$

which completes the proof.

Note. $n = 0$ implies the trivial case $z^0 = r^0 \operatorname{cis} 0 = 1$.

10. Th. 5.1.3.6 holds also for fractional values of n .

PROOF:

Let

$$w = z^{1/n} = \sqrt[n]{z} \quad (1)$$

so that w is a solution of

$$w^n = z \quad (2)$$

Then, by Df. 5.1.3.5,

$$w = R \operatorname{cis} \varphi \quad \text{and} \quad z = r \operatorname{cis} \theta \quad (3)$$

From (2), (3), and Prob. 9 above,

$$w^n = R^n \operatorname{cis} (n\varphi) = r \operatorname{cis} \theta$$

which implies

$$R^n = r \quad \text{and} \quad n\varphi = \theta \pm 2k\pi, \quad \text{where } k = 0, 1, 2, \dots$$

i.e.

$$R = \sqrt[n]{r} \quad \text{and} \quad \varphi = (\theta \pm 2k\pi)/n, \quad k = 0, 1, 2, \dots$$

which then entail, by (3),

$$w = \sqrt[n]{r} \operatorname{cis} ((\theta \pm 2k\pi)/n)$$

and, by (1),

$$z^{1/n} = (r \operatorname{cis} \theta)^{1/n} = r^{1/n} \operatorname{cis} ((\theta \pm 2k\pi)/n) \quad (4)$$

where $\cos((\theta \pm 2k\pi)/n)$ and $\sin((\theta \pm 2k\pi)/n)$ have the same values for two integers k differing by a multiple of n , since they are of course periodic. Hence (4) yields exactly n distinct values for $z^{1/n}$, viz.,

$$\sqrt[n]{z} = r^{1/n} \operatorname{cis} ((\theta \pm 2k\pi)/n), \quad k = 0, 1, 2, \dots, n-1 \quad (5)$$

Hence, by Prob. 9 above and (5),

$$z^{m/n} = r^{m/n} \operatorname{cis} (m(\theta \pm 2k\pi)/n) \quad (6)$$

for every $m, n \in I$, which completes the proof.

11. Find all roots of (i) $\sqrt[n]{1}$ and (ii) $\sqrt[3]{1+i}$.

Solution:

(i) By Prob. 10, $1^{1/n} = 1$, $\theta = 0$, and

$$\sqrt[n]{1} = \operatorname{cis} (2k\pi/n), \quad k = 0, 1, 2, \dots, n-1$$

which exhausts the n roots of unity. (Pictorially, they coincide with the vertices of a regular polygon of n sides inscribed in the unit circle, with one vertex of the polygon at $z = 1$.)

(ii) Since $1 + i = \sqrt{2} \operatorname{cis} (\pi/4)$, it follows from Prob. 10, (5), that

$$\sqrt[3]{1+i} = \sqrt[6]{2} \operatorname{cis} (((\pi/4) + 2k\pi)/3), \quad k = 0, 1, 2$$

Hence the three roots are:

$$w_1 = 2^{1/6} \operatorname{cis} (\pi/12), \quad w_2 = 2^{1/6} \operatorname{cis} (3\pi/4), \quad w_3 = 2^{1/6} \operatorname{cis} (17\pi/12)$$

12. Prove Th. 5.1.3.7.

PROOF:

Let $z_1 = a + ib \neq 0$ and $z_2 = c + id \neq 0$, and also let O denote the origin. Then, if O, z_1 , and z_2 are not collinear, O, z_1, z_2 , and $z_1 + z_2$ are the vertices of a parallelogram (cf. Fig. 5.1.3.b), where $|z_1|$ and $|z_2|$ are two sides of the parallelogram and $|z_1 + z_2|$ is a diagonal. That is, $|z_1|, |z_2|$, and $|z_1 + z_2|$ are three sides of a triangle. Hence

$$|z_1 + z_2| < |z_1| + |z_2| \quad (1)$$

and iff $\arg z_1 = \arg z_2$,

$$|z_1 + z_2| = |z_1| + |z_2| \quad (2)$$

Combining (1) and (2),

$$|z_1 + z_2| \leq |z_1| + |z_2| \quad (3)$$

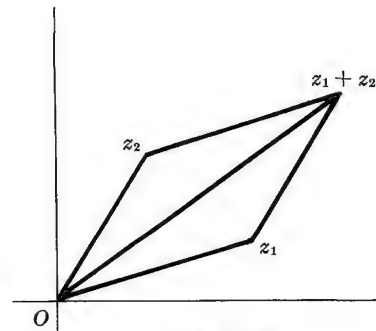


Fig. 5.1.3b

which immediately implies $|z_1 + (-z_2)| \leq |z_1| + |(-z_2)| = |z_1| + |z_2|$, i.e.

$$|z_1 - z_2| \leq |z_1| + |z_2| \quad (4)$$

Furthermore, from (3), $|z_1| = |z_1 - z_2 + z_2| \leq |z_1 - z_2| + |z_2|$, i.e.

$$|z_1| - |z_2| \leq |z_1 - z_2| \quad (5)$$

Combining (4) and (5),

$$|z_1| - |z_2| \leq |z_1 - z_2| \leq |z_1| + |z_2|$$

which completes the proof.

If either $z_1 = 0$ or $z_2 = 0$ or if $z_1 = z_2 = 0$, the case is obviously trivial, yielding the same result.

Second Proof. A direct computation yields

$$\sqrt{(a+c)^2 + (b+d)^2} \leq \sqrt{a^2 + b^2} + \sqrt{c^2 + d^2}$$

which is evidently true, since, by squaring both sides of the inequality,

$$(a+c)^2 + (b+d)^2 \leq a^2 + b^2 + c^2 + d^2 + 2\sqrt{(a^2 + b^2)(c^2 + d^2)}$$

i.e.

$$ac + bd \leq \sqrt{(a^2 + b^2)(c^2 + d^2)}$$

and, squaring again,

$$a^2c^2 + 2abcd + b^2d^2 \leq a^2c^2 + a^2d^2 + b^2c^2 + b^2d^2$$

i.e.

$$(ad - bc)^2 \geq 0$$

which of course holds, since $a, b, c, d \in \bar{R}$. Hence

$$|z_1 + z_2| \leq |z_1| + |z_2|$$

The rest can be proved likewise.

13. If $x, y \in C$, then

- (i) $\overline{x+y} = \bar{x} + \bar{y}$, (ii) $\overline{x-y} = \bar{x} - \bar{y}$, (iii) $\overline{x \cdot y} = \bar{x} \cdot \bar{y}$, (iv) $\overline{x/y} = \bar{x}/\bar{y}$, where $y \neq 0$

PROOF:

Let $x = a + ib$ and $y = c + id$; then

- (i) $\overline{x+y} = \overline{(a+c) + i(b+d)} = (a+c) - i(b+d) = (a-ib) + (c-id) = \bar{x} + \bar{y}$.
(ii) Likewise $\overline{x-y} = \bar{x} - \bar{y}$.
(iii) $\overline{x \cdot y} = \overline{(ac-bd) + i(ad+bc)} = (ac-bd) - i(ad+bc) = (a-ib)(c-id) = \bar{x} \cdot \bar{y}$.
(iv) Likewise $\overline{x/y} = \bar{x}/\bar{y}$. (Or, directly from (iii), it follows that $\bar{y} \cdot (x/y) = \overline{y \cdot (x/y)} = \bar{x}$, and that, dividing both sides by $\bar{y} \neq 0$, $x/y = \bar{x}/\bar{y}$.)

14. If $x, y \in C$, then

- (i) $|x| = |\bar{x}|$, (ii) $x\bar{x} = |x|^2$, (iii) $|x \cdot y| = |x| \cdot |y|$, (iv) $|x/y| = |x|/|y|$, where $y \neq 0$

PROOF:

- (i) Let $x = a + ib$; then $\bar{x} = a - ib$ and, since $|x| = (a^2 + b^2)^{1/2}$ and $|\bar{x}| = (a^2 + (-b)^2)^{1/2}$, it follows at once that $|x| = |\bar{x}|$.
(ii) $x\bar{x} = (a+ib)(a-ib) = a^2 + b^2 = ((a^2 + b^2)^{1/2})^2 = |x|^2$
(iii) $|x \cdot y| = |(ac-bd) + i(ad+bc)| = ((ac-bd)^2 + (ad+bc)^2)^{1/2}$
 $= (a^2c^2 + a^2d^2 + b^2c^2 + b^2d^2)^{1/2} = (a^2 + b^2)^{1/2} (c^2 + d^2)^{1/2}$
 $= |x| \cdot |y|$

Second Proof. By (ii) immediately above, $|x \cdot y|^2 = (x \cdot y)(\overline{x \cdot y}) = x \cdot y \cdot \bar{x} \cdot \bar{y} = x \cdot \bar{x} \cdot y \cdot \bar{y} = |x|^2 \cdot |y|^2$, and since $|x \cdot y| \geq 0$, $|x| \geq 0$, and $|y| \geq 0$, it follows that $|x \cdot y| = |x| \cdot |y|$.

Note. $|x_1 x_2 \cdots x_n| = |x_1| |x_2| \cdots |x_n|$, for every $x_1, x_2, \dots, x_n \in C$, evidently holds as a generalized result of (iii).

- (iv) $|x/y| = |(ac+bd) + i(ad-bc)| / (c^2 + d^2) = \sqrt{(ac+bd)^2 + (ad-bc)^2} / (c^2 + d^2)$
 $= \sqrt{(a^2 + b^2)(c^2 + d^2)} / (c^2 + d^2) = \sqrt{a^2 + b^2} / \sqrt{c^2 + d^2} = |x| / |y|$

Second Proof. By (iii), $|y| |x/y| = |y(x/y)| = |x|$, and, dividing both sides by $|y| \neq 0$, $|x/y| = |x| / |y|$.

15. If $|z_1| = |z_2| = \dots = |z_n| = 1$, then

$$|z_1 + z_2 + \dots + z_n| = |1/z_1 + 1/z_2 + \dots + 1/z_n|$$

PROOF:

Since, by Prob. 14, (i), $|z_1| = |\bar{z}_1|$ and evidently also

$$|z_1 + z_2| = |\bar{z}_1 + \bar{z}_2|$$

(or $z_1 = a_1 + ib_1$ and $z_2 = a_2 + ib_2$ imply $\bar{z}_1 = a_1 - ib_1$ and $\bar{z}_2 = a_2 - ib_2$, which in turn imply $z_1 + z_2 = (a_1 + a_2) + i(b_1 + b_2)$ and $\bar{z}_1 + \bar{z}_2 = (a_1 + a_2) - i(b_1 + b_2)$, i.e. $|z_1 + z_2| = ((a_1 + a_2)^2 + (b_1 + b_2)^2)^{1/2} = |\bar{z}_1 + \bar{z}_2|$), this result can be readily generalized to

$$|z_1 + z_2 + \dots + z_n| = |\bar{z}_1 + \bar{z}_2 + \dots + \bar{z}_n|$$

Since, by Prob. 14, (ii), and the initial hypothesis, $|z_1|^2 = z_1 \bar{z}_1 = 1$, it follows at once that $\bar{z}_1 = 1/z_1$. Likewise $\bar{z}_2 = 1/z_2, \dots, \bar{z}_n = 1/z_n$. Hence

$$|z_1 + z_2 + \dots + z_n| = |1/z_1 + 1/z_2 + \dots + 1/z_n|$$

16. If $x, y \in C$, and if $|x| = 1$ or $|y| = 1$, then

$$|(x + y)/(1 + \bar{x}y)| = 1$$

PROOF:

Let $|x| = 1$; then, since $x\bar{x} = |x|^2 = 1$, i.e. $x = 1/\bar{x}$, it follows that

$$x + y = (1/\bar{x}) + y = (1 + \bar{x}y)/x$$

$$\therefore (x + y)/(1 + \bar{x}y) = ((1 + \bar{x}y)/\bar{x})/(1 + \bar{x}y) = 1/\bar{x}$$

$$\therefore |(x + y)/(1 + \bar{x}y)| = |1/\bar{x}| = 1/|\bar{x}| = 1/|x| = 1$$

The same result is obtained likewise by assuming $|y| = 1$, which was a part of the initial hypothesis.

17. If $z \in C$, where $z = a + ib$, then (i) $\bar{\bar{z}} = z$, (ii) $z + \bar{z} = 2 \operatorname{Re}(z)$, (iii) $z - \bar{z} = 2 \operatorname{Im}(z)$, where $\operatorname{Re}(z)$ and $\operatorname{Im}(z)$ are the real and imaginary part of z , respectively (cf. Df. 5.1.3.1).

PROOF:

(i) $\bar{z} = a - ib$, by hypothesis and Df. 5.1.3.8 itself. Hence $\bar{\bar{z}} = \overline{(a - ib)} = a + ib = z$, i.e. $\bar{\bar{z}} = z$.

(ii) $z + \bar{z} = (a + ib) + (a - ib) = 2a = 2 \operatorname{Re}(z)$

(iii) $z - \bar{z} = (a + ib) - (a - ib) = 2ib = 2 \operatorname{Im}(z)$

18. If $x, y \in C$, then (i) $|x + y|^2 = |x|^2 + |y|^2 + 2 \operatorname{Re}(x\bar{y})$, (ii) $|x - y|^2 = |x|^2 + |y|^2 - 2 \operatorname{Re}(x\bar{y})$, (iii) $|x + y|^2 + |x - y|^2 = 2(|x|^2 + |y|^2)$, where $\operatorname{Re}(x\bar{y})$ denotes the real part of $x\bar{y}$ (cf. Df. 5.1.3.1).

PROOF:

(i) By Prob. 13, (i), and Prob. 14, (ii),

$$|x + y|^2 = (x + y)\overline{(x + y)} = (x + y)(\bar{x} + \bar{y}) = x\bar{x} + y\bar{y} + x\bar{y} + \bar{x}y$$

where, by Prob. 14, (ii), $x\bar{x} = |x|^2$, $y\bar{y} = |y|^2$, and, by Prob. 13, (iii), Prob. 17, (i), and Df. 5.1.3.8, $\bar{x}y = \bar{x} \cdot \bar{\bar{y}} = x\bar{\bar{y}}$, which implies, by Prob. 17, (iii), $\bar{x}y + x\bar{y} = x\bar{\bar{y}} + x\bar{y} = 2 \operatorname{Re}(x\bar{y})$. Hence

$$|x + y|^2 = |x|^2 + |y|^2 + 2 \operatorname{Re}(x\bar{y}) \quad (1)$$

(ii) Likewise, going through similar steps as above,

$$|x - y|^2 = |x|^2 + |y|^2 - 2 \operatorname{Re}(x\bar{y}) \quad (2)$$

(iii) Adding (1) and (2),

$$|x + y|^2 + |x - y|^2 = 2(|x|^2 + |y|^2) \quad (3)$$

which completes the proof.

19. For every $a_k, b_k \in C$, $k = 1, 2, \dots, n$,

$$|\sum_k a_k b_k|^2 \leq (\sum_k |a_k|^2)(\sum_k |b_k|^2)$$

PROOF:

Let $d \in C$ such that

$$a_k = d\bar{b}_k + c_k, \quad k = 1, 2, \dots, n \quad (1)$$

Then $|a_k|^2 = a_k \bar{a}_k = (d\bar{b}_k + c_k)(d\bar{b}_k + \bar{c}_k) = |d|^2 |b_k|^2 + |c_k|^2 + d\bar{b}_k c_k + d\bar{b}_k \bar{c}_k$. Hence

$$\sum_k |a_k|^2 = |d|^2 \sum_k |b_k|^2 + \sum_k |c_k|^2 + d \sum_k \bar{b}_k c_k + d \overline{(\sum_k b_k c_k)} \quad (2)$$

and, from (1) itself,

$$\sum_k a_k b_k = d \sum_k \bar{b}_k b_k + \sum_k c_k b_k = d \sum_k |b_k|^2 + \sum_k b_k c_k \quad (3)$$

If every b_k is not to vanish all simultaneously, and if d is to satisfy

$$\sum_k a_k b_k = d \sum_k |b_k|^2 \quad (4)$$

then, by (3), $\sum_k b_k c_k = 0$.

Hence, by (2),

$$\sum_k |a_k|^2 = |d|^2 \sum_k |b_k|^2 + \sum_k |c_k|^2$$

Multiplying both sides of the equation by $\sum_k |b_k|^2$,

$$(\sum_k |a_k|^2)(\sum_k |b_k|^2) = |d|^2 (\sum_k |b_k|^2)^2 + (\sum_k |c_k|^2)(\sum_k |b_k|^2) \geq (|d| \sum_k |b_k|^2)^2 = |\sum_k a_k b_k|^2$$

by (4). The equality holds evidently iff

$$(\sum_k |c_k|^2)(\sum_k |b_k|^2) = 0$$

i.e. either $\sum_k |b_k|^2 = 0$ which implies $b_1 = b_2 = \dots = b_n = 0$, or $\sum_k |c_k|^2 = 0$ which implies $c_1 = c_2 = \dots = c_n = 0$. (The latter also implies, by (1), $a_k = d\bar{b}_k$.)

Second Proof. If the inequality has already been proved to hold with respect to real numbers (cf. §5.1.2, Prob. 13), i.e.,

$$(\sum_k |a_k| |b_k|)^2 \leq (\sum_k |a_k|^2)(\sum_k |b_k|^2), \quad k = 1, 2, \dots, n \quad (1)$$

then, since

$$|\sum_k a_k b_k| \leq \sum_k |a_k| |b_k|, \quad \text{i.e.} \quad |\sum_k a_k b_k|^2 \leq (\sum_k |a_k| |b_k|)^2 \quad (2)$$

it immediately follows from (1), (2), and transitivity that

$$|\sum_k a_k b_k|^2 \leq (\sum_k |a_k|^2)(\sum_k |b_k|^2)$$

which completes the proof.

20. If $|a| < 1$, then $|(z-a)/(1-\bar{a}z)| < 1$, $= 1$, > 1 according as $|z| < 1$, $= 1$, > 1 .

PROOF:

Since

$$\begin{aligned} |z-a|^2 - |1-\bar{a}z|^2 &= (z-a)(\bar{z}-\bar{a}) - (1-\bar{a}z)(1-\overline{a\bar{z}}) = (z-a)(\bar{z}-\bar{a}) - (1-\bar{a}z)(1-a\bar{z}) \\ &= (z\bar{z}-1)(1-a\bar{a}) = (|z|^2-1)(1-|a|^2) \end{aligned}$$

and since, by hypothesis, $|a| < 1$, i.e. $1-|a|^2 > 0$, it follows that $|z-a|^2 - |1-\bar{a}z|^2 < 1$, $= 1$, > 1 according as $|z|^2 < 1$, $= 1$, > 1 .

Hence $|z-a|^2 \leq |1-\bar{a}z|^2$, i.e. $|(z-a)/(1-\bar{a}z)| \leq 1$ according as $|z| \leq 1$, completing the proof.

21. Two numbers x and y , whose sum as well as product is real, are either both real numbers or conjugate complex numbers.

PROOF:

Let $a, b \in \bar{R}$; then any two numbers x and y , where $x + y = a$ and $xy = b$, are in fact the roots of the following quadratic equation

$$t^2 - at + b = 0$$

where the roots are either both real numbers if $a^2 - 4b \geq 0$ or conjugate complex numbers if $a^2 - 4b < 0$, i.e.

$$x = (a/2) + i(\sqrt{4b - a^2}/2) \quad \text{and} \quad y = (a/2) - i(\sqrt{4b - a^2}/2)$$

Second Proof. Let $x = a + ib$ and $y = c + id$; then

$$x + y = (a + c) + i(b + d) \quad \text{and} \quad xy = (ac - bd) + i(ad + bc)$$

which by hypothesis implies both $b + d = 0$ and $ad + bc = 0$, from which it follows that $b = -d$ and $b(a - c) = 0$. Hence either $b = 0$ or $a = c$ if $b \neq 0$, which implies that either $b = d = 0$ or $a = c$ and $b = -d$. If $b = d = 0$, then evidently $x = a$ and $y = c$, i.e. $x, y \in \bar{R}$, and if $a = c$ and $b = -d$, then $x, y \in C$ and $x = \bar{y}$ or $y = \bar{x}$, which completes the proof.

22. If a is an imaginary root of the equation: $z^n - 1 = 0$, then

$$1 + a + a^2 + \dots + a^{n-1} = 0$$

PROOF:

Since $z^n - 1 = (z - 1)(z^{n-1} + z^{n-2} + \dots + z + 1)$, and since a is a root of the equation and evidently $a - 1 \neq 0$,

$$(a - 1)(a^{n-1} + a^{n-2} + \dots + a + 1) = 0$$

implies at once that

$$a^{n-1} + a^{n-2} + \dots + a^2 + a + 1 = 0$$

which completes the proof.

23. If $r_k \in \bar{R}$ and $c_k \in C$, $k = 0, 1, \dots, n$, then (i) $\overline{\sum r_k c_k} = \sum r_k \bar{c}_k$, (ii) $\bar{c}^n = (\bar{c})^n$ for every $c \in C$ and $n \in N$.

PROOF:

(i) Generalizing Prob. 13, (i),

$$\overline{r_1 c_1 + r_2 c_2 + \dots + r_n c_n} = \overline{r_1 c_1} + \overline{r_2 c_2} + \dots + \overline{r_n c_n}$$

But, by Prob. 13, (iii),

$$\overline{r_1 c_1} = \bar{r}_1 \bar{c}_1, \quad \overline{r_2 c_2} = \bar{r}_2 \bar{c}_2, \quad \dots, \quad \overline{r_n c_n} = \bar{r}_n \bar{c}_n$$

where $\bar{r}_k = r_k$, since $r_k \in \bar{R}$. Hence

$$\overline{\sum r_k c_k} = \sum r_k \bar{c}_k$$

(ii) Let $x = y$ in Prob. 13, (iii); then $\overline{x^2} = (\bar{x})^2$, and likewise $\overline{x^3} = \overline{x^2 x} = \overline{x^2} \cdot \bar{x} = (\bar{x})^2 \bar{x} = (\bar{x})^3$. Since both $x \in C$ and $c \in C$, x and c are interchangeable; hence, by induction,

$$\bar{c}^n = (\bar{c})^n$$

which completes the proof.

24. If $f(x) = \sum_k a_k x^k$, $k=0,1,\dots,n$, where $a_k \in \bar{R}$, then

$$\overline{f(z)} = f(\bar{z})$$

for every $z \in C$.

PROOF:

Since $f(z) = \sum_k a_k z^k$, it follows, by Prob. 23, (i), (ii), that

$$\begin{aligned} \overline{f(z)} &= \overline{a_0 z^0 + a_1 z^1 + \dots + a_n z^n} = \overline{a_0 z^0} + \overline{a_1 z^1} + \dots + \overline{a_n z^n} \\ &= a_0 (\bar{z})^0 + a_1 (\bar{z})^1 + \dots + a_n (\bar{z})^n = f(\bar{z}) \end{aligned}$$

which completes the proof.

25. If, as in Prob. 24 above, $f(x) = \sum_k a_k x^k$, $k=0,1,\dots,n$, where $a_k \in \bar{R}$, and if $f(z) = a + ib$, where $a, b \in \bar{R}$, then $f(\bar{z}) = a - ib$.

PROOF:

Since $f(z) = a + ib$, it follows at once, by Df. 5.1.3.8, that $\overline{f(z)} = \overline{a + ib} = a - ib$. But, by Prob. 24, $\overline{f(z)} = f(\bar{z})$. Hence

$$f(\bar{z}) = a - ib$$

26. If a complex number z is a root of the equation

$$f(x) = \sum_k a_k x^k = 0 \tag{1}$$

where $a_k \in \bar{R}$, then \bar{z} is also a root of (1).

PROOF:

Since z is a root of (1), by hypothesis, it follows from Prob. 25 above, that

$$f(z) = a + ib = 0 \tag{2}$$

where $a, b \in \bar{R}$. Since the real part and the imaginary part are linearly independent, (2) implies that $a = b = 0$, which in turn implies, together with Prob. 25,

$$f(\bar{z}) = a - ib = 0$$

completing the proof.

27. If $u \in C$ is a root of the equation:

$$f(z) = \sum_k c_k z^k = 0, \quad k=0,1,\dots,n \tag{1}$$

where z is the indeterminate of (1) and $c_k \in C$, then \bar{u} is a root of the equation

$$g(z) = \sum_k \bar{c}_k z^k = 0$$

PROOF:

Since $f(u) = 0$, it follows from Prob. 26 that $f(\bar{u}) = 0$, which in turn implies, by Prob. 24,

$$f(\bar{u}) = \overline{f(u)} = 0$$

$$\text{i.e.} \quad \overline{f(u)} = \overline{c_0 u^0 + c_1 u^1 + \dots + c_n u^n} = \bar{c}_0 \bar{u}^0 + \bar{c}_1 \bar{u}^1 + \dots + \bar{c}_n \bar{u}^n = 0$$

Hence, by substitution [and Prob. 17, (i)],

$$f(\bar{u}) = \bar{c}_0 \bar{u}^0 + \bar{c}_1 \bar{u}^1 + \dots + \bar{c}_n \bar{u}^n = \bar{c}_0 u^0 + \bar{c}_1 u^1 + \dots + \bar{c}_n u^n = g(\bar{u}) = 0$$

which proves that \bar{u} is a root of $g(z) = 0$.

28. Find the roots of the equation $z^2 = a + ib$, where $a, b \in \bar{R}$ and $b \neq 0$.

Solution:

Let $z = x + iy$; then, by hypothesis,

$$(x + iy)^2 = a + ib, \quad \text{i.e.} \quad x^2 - y^2 + i2xy = a + ib \tag{1}$$

which, by Prob. 1, implies

$$(i) \quad x^2 - y^2 = a, \quad (ii) \quad 2xy = b \quad (2)$$

Squaring (i) and (ii), then adding them,

$$(x^2 + y^2)^2 = a^2 + b^2, \quad \text{i.e.} \quad x^2 + y^2 = (a^2 + b^2)^{1/2} \quad (3)$$

From (3) and (2), (i),

$$x^2 = ((a^2 + b^2)^{1/2} + a)/2, \quad y^2 = ((a^2 + b^2)^{1/2} - a)/2 \quad (4)$$

$$\text{i.e.} \quad x = \pm(((a^2 + b^2)^{1/2} + a)/2)^{1/2}, \quad y = \pm(((a^2 + b^2)^{1/2} - a)/2)^{1/2} \quad (5)$$

since the right sides of the equations of (4) are both non negative.

In general, since x and y must satisfy (2), (ii), as well as (3) and (2), (i), xy and b must have the same sign if $b \neq 0$. Hence the square roots of $a + ib$ are

$$(i) \quad p + iq \quad \text{and} \quad -p - iq \quad \text{if} \quad b > 0,$$

$$(ii) \quad p - iq \quad \text{and} \quad -p + iq \quad \text{if} \quad b < 0,$$

$$\text{where} \quad p = (((a^2 + b^2)^{1/2} + a)/2)^{1/2} \quad \text{and} \quad q = (((a^2 + b^2)^{1/2} - a)/2)^{1/2}.$$

Note. $b = 0$ implies that

$$(i) \quad x = y = 0 \quad \text{if} \quad a = 0$$

$$(ii) \quad z = \pm\sqrt{a} \quad \text{if} \quad a > 0, \quad \text{since} \quad (a^2 + b^2)^{1/2} = a, \quad \text{implying by (4) that} \quad x = \pm\sqrt{a} \quad \text{and} \quad y = 0.$$

$$(iii) \quad z = \pm i\sqrt{-a} \quad \text{if} \quad a < 0, \quad \text{since} \quad (a^2 + b^2)^{1/2} = -a, \quad \text{implying by (4) that} \quad x = 0 \quad \text{and} \quad y = \pm\sqrt{-a}.$$

29. Find the roots of the quadratic equation $rz^2 + 2sz + t = 0$, where $r \neq 0$ and $r, s, t \in C$.

Solution:

Let $s' = s/r$ and $t' = t/r$; then, by hypothesis,

$$z^2 + 2s'z + t' = 0, \quad \text{i.e.} \quad (z + s')^2 = s'^2 - t' \quad (1)$$

which implies, by substituting $z = x + iy$, $s' = b + ib'$, $t' = c + ic'$,

$$\begin{aligned} ((x + iy) + (b + ib'))^2 &= (b + ib')^2 - (c + ic') \\ \text{i.e.} \quad ((x + b) + i(y + b'))^2 &= (b^2 - b'^2 - c) + i(2bb' - c') \end{aligned}$$

Then, introducing two sets of substitutions:

$$x + b = u, \quad y + b' = v \quad (2)$$

$$b^2 - b'^2 - c = p, \quad 2bb' - c' = q \quad (3)$$

it follows that

$$(u + iv)^2 = p + iq \quad (4)$$

which reduces the problem to Prob. 28, since (4) is of the form of Prob. 28. Hence, taking similar steps as in Prob. 28,

$$\begin{aligned} u &= \pm(((p^2 + q^2)^{1/2} + p)/2)^{1/2}, & v &= \pm(((p^2 + q^2)^{1/2} - p)/2)^{1/2} \\ \text{and} \quad x &= b \pm (((p^2 + q^2)^{1/2} + p)/2)^{1/2}, & y &= b' \pm (((p^2 + q^2)^{1/2} - p)/2)^{1/2} \end{aligned}$$

30. The roots of the quadratic equation: $z^2 + 2s + t = 0$ are not conjugate to each other if $s \notin \bar{R}$ or $t \notin \bar{R}$.

PROOF:

Suppose the given equation has the roots z_1 and z_2 which are conjugate to each other. Then, by Prob. 17, (ii), and Prob. 14, (ii),

$$z_1 + z_2 = 2R(z_1) = 2R(z_2), \quad \text{and} \quad z_1 \cdot z_2 = |z_1|^2 = |z_2|^2$$

On the other hand, since $z_1 + z_2 = -2s$ and $z_1 \cdot z_2 = t$ (cf. Prob. 21 above), it must be the case that $-2s = 2R(z_1)$ and $t = |z_1|^2$, i.e. $s \in \bar{R}$ and $t \in \bar{R}$, contradictory to the initial hypothesis.

Hence neither $z_1 = \bar{z}_2$ nor $z_2 = \bar{z}_1$.

31. The quadratic equation with complex coefficients:

$$z^2 + 2(a + ia')z + b + ib' = 0 \quad (1)$$

where $a \neq 0$ and $a' \neq 0$, has at least one real root iff

$$b'^2 - 4aa'b' + 4a'^2c = 0 \quad (2)$$

and has pure imaginary roots iff

$$-b'^2 + 4aa'c' + 4a^2c = 0 \quad (3)$$

PROOF:

(i) Let $z = r$, where $r \in \bar{R}$; then, substituting r in (1),

$$r^2 + 2(a + ia')r + b + ib' = 0$$

$$\text{i.e. } (r^2 + 2ar + b) + i(2a'r + b') = 0.$$

Hence, by Prob. 1, it follows that

$$r^2 + 2ar + b = 0 \quad \text{and} \quad 2a'r + b' = 0$$

from which r is cancelled to yield (2). The converse also holds, as can be verified by taking the steps in reverse.

(ii) Let $z = iu$, where $u \in \bar{R}$; then, substituting iu in (1) and taking similar steps as in (i) above, (3) is duly obtained. The converse holds likewise.

32. Prove Th. 5.1.3.9.

PROOF:

Let $z_1, z_2 \in C$, and $z_1 \leftrightarrow \bar{z}_1$, $z_2 \leftrightarrow \bar{z}_2$, where $\bar{z}_1 = V(z_1)$ and $\bar{z}_2 = V(z_2)$. Then, by Prob. 13, (i), (iii),

$$z_1 + z_2 \leftrightarrow V(z_1 + z_2) = V(z_1) + V(z_2) \leftrightarrow \bar{z}_1 + \bar{z}_2$$

$$z_1 \cdot z_2 \leftrightarrow V(z_1 \cdot z_2) = V(z_1) \cdot V(z_2) \leftrightarrow \bar{z}_1 \cdot \bar{z}_2$$

which implies that V is an isomorphism. Furthermore, since $\bar{z}_1, \bar{z}_2 \in C$ by Df. 5.1.3.8 itself, V is an automorphism, completing the proof.

33. Prove Th. 5.1.3.10.

PROOF:

Since $z \in C$ implies $\bar{\bar{z}} = z$, by Prob. 17, (i), it follows that $z_1, z_2 \in C$ implies, by Prob. 13, (i),

$$z_1 + z_2 \leftrightarrow V^2(z_1 + z_2) = \overline{z_1 + z_2} = \bar{\bar{z}}_1 + \bar{\bar{z}}_2 = V^2(z_1) + V^2(z_2) \leftrightarrow z_1 + z_2$$

and likewise, by Prob. 13, (iii), and Prob. 17, (i),

$$z_1 \cdot z_2 \leftrightarrow V^2(z_1 \cdot z_2) = V^2(z_1) \cdot V^2(z_2) \leftrightarrow z_1 \cdot z_2$$

Hence V^2 is evidently an involution.

34. If $ad - bc \neq 0$, where $a, b, c, d \in C$, then the linear transformation f (cf. Df. 5.1.3.11):

$$w = f(z) = (az + b)/(cz + d) \quad (1)$$

defines the inverse transformation f^{-1} of f , viz.

$$z = f^{-1}(w) = (-dw + b)/(cw - a) \quad (2)$$

PROOF:

Solve (1) for z , and (2) immediately follows. Furthermore, since the coefficients of (2) are obtained by merely changing a and d of (1) into $-d$ and $-a$ respectively, the original condition remains valid throughout the transformation, viz.

$$(-d)(-a) - bc = ad - bc \neq 0$$

$$\begin{aligned} \text{Note. } f^{-1}f &= f^{-1}((az + b)/(cz + d)) \\ &= (a((-dw + b)/(cw - a)) + b)/(c((-dw + b)/(cw - a)) + d) \\ &= w. \end{aligned}$$

$$\begin{aligned}
 \text{Likewise } ff^{-1} &= f((-dw + b)/(cw - a)) \\
 &= (-d((az + b)/(cz + d)) + b)/(c((a + b)/(cz + d)) - a) \\
 &= z.
 \end{aligned}$$

Hence the identity transformation I is defined by $w = z$, viz.

$$w = I(z) = z \quad \text{or} \quad z = I(w) = w$$

which is a special case of the linear transformation with $a = d \neq 0$, $b = c = 0$.

35. If $a + d = 0$ for $w = f(z) = (az + b)/(cz + d)$, then $z = f(w)$.

PROOF:

Since $w = (az + b)/(cz + d)$, by hypothesis, i.e.

$$czw + dw - az - b = 0 \tag{1}$$

the left-hand side of the equation (1) is symmetric with respect to z and w if $a + d = 0$.

Hence $w = f(z)$ implies $z = f(w)$.

Second Proof. Solve (1) for z , substituting $a = -d$ and $d = -a$, and $z = (b - dw)/(cw - a) = (aw + b)/(cw + d) = f(w)$.

36. The Möbius transformation $w = f(z)$ such that $f(f(z)) = z$, is exhausted by two, and only two, functions: $w = z$ and $w = (az + b)/(cz - a)$.

PROOF:

Since, by Df. 5.1.3.11, $f(z) = (az + b)/(cz + d)$, it follows that, by direct substitution,

$$f(f(z)) = (a((az + b)/(cz + d)) + b)/(c((az + b)/(cz + d)) + d) = z$$

by hypothesis. Simplify the equation, and

$$(a + d)(cz^2 - (a - d)z - b) = 0$$

which implies $a + d = 0$ or $cz^2 - (a - d)z - b = 0$. If $a + d = 0$, i.e. $d = -a$, then

$$w = f(z) = (az + b)/(cz - a)$$

and if $cz^2 - (a - d)z - b = 0$, then $c = a - d = b = 0$, i.e.

$$w = f(z) = z$$

completing the proof.

37. If z is linearly mapped into w which in turn is linearly mapped into w' , then the double mapping is also linear.

PROOF:

Let

$$w = f(z) = (az + b)/(cz + d), \quad \text{where } ad - bc \neq 0$$

and

$$w' = g(w) = (a'w + b')/(c'w + d'), \quad \text{where } a'd' - b'c' \neq 0$$

Then, by substitution,

$$\begin{aligned}
 w' = g(f(z)) &= (a'((az + b)/(cz + d)) + b')/(c'((az + b)/(cz + d)) + d') \\
 &= ((a'a + cb')z + (ba' + db'))/((ac' + cd')z + (bc' + dd'))
 \end{aligned}$$

which is evidently of the form of linear transformations, where also

$$(aa' + cb')(bc' + dd') - (ba' + db')(ac' + cd') = (ad - bc)(a'd' - b'c') \neq 0$$

Hence the outcome of successive linear (i.e. Möbius) transformations is again linear.

Note. In this context, $ad - bc \neq 0$ and $a'd' - b'c' \neq 0$ may be called the *determinants* of the mappings f and g respectively. The determinant of the double transformations, i.e. the *product*, of f and g , is also the product of the determinants of f and g , as has already been shown above.

38. Möbius transformations are not commutative.

PROOF:

Let the mapping of f and g be linear with respect to z , i.e.

$$z_1 = f(z) = (az + b)/(cz + d), \quad \text{where } ad - bc \neq 0$$

and

$$z_2 = g(z_1) = (a'z_1 + b')/(c'z_1 + d'), \quad \text{where } a'd' - b'c' \neq 0$$

Then, as in Prob. 27 above,

$$z_2 = g(f(z)) = ((aa' + cb')z + (ba' + db'))/((ac' + cd')z + (bc' + dd')) \quad (1)$$

where $(ad - bc)(a'd' - b'c') \neq 0$, while

$$z' = g(z) = (a'z + b')/(c'z + d'), \quad \text{where } a'd' - b'c' \neq 0$$

and

$$z'' = f(z') = (az' + b)/(cz' + d), \quad \text{where } ad - bc \neq 0$$

imply similarly

$$z'' = f(g(z)) = ((a'a + c'b)z + (b'a + d'b))/((a'c + c'd)z + (b'c + d'd)) \quad (2)$$

Hence, in general,

$$z_2 = g(f(z)) \neq f(g(z)) = z''$$

i.e. Möbius transformations are not commutative.

Note. a, b, c, d and a', b', c', d' are interchanged in (1) and (2), which implies that the transformations are commutative iff $a = a', b = b', c = c', d = d'$, which is rather trivial.

39. Möbius transformations are associative.

PROOF:

Let, as in Prob. 38,

$$z_1 = f(z) = (az + b)/(cz + d), \quad \text{where } ad - bc \neq 0$$

$$z_2 = g(z_1) = (a'z_1 + b')/(c'z_1 + d'), \quad \text{where } a'd' - b'c' \neq 0$$

$$z_3 = h(z_2) = (a''z_2 + b'')/(c''z_2 + d''), \quad \text{where } a''d'' - b''c'' \neq 0$$

Then, by Prob. 37, either

$$z_3 = h(z_2) = h(g(z_1)) = h(g(f(z))) = h(gf(z))$$

or

$$z_3 = h(z_2) = h(g(z_1)) = (hg)(z_1) = (hg)f(z)$$

and in either case, $z_3 = hgf(z)$, i.e.

$$z_3 = ((a''(aa' + cb') + b''(ac' + cd'))z + (a''(ba' + db')))/((c''(aa' + cb') + d''(ac' + cd'))z + (c''(ba' + db') + d''(bc' + dd')))$$

where $(ad - bc)(a'd' - b'c')(a''d'' - b''c'') \neq 0$, which establishes the associativity: $h(gf) = (hg)f$ in Möbius transformation.

40. Prove Th. 5.1.3.12.

PROOF:

The closure property (G1) has already been established by Prob. 37, and the associativity (G2) by Prob. 39 above. As for the identity (G3), it is defined by the transformation $I: w = f(z) = z$ (cf. Prob. 34 note), and the inverse (G4) is obtained by Prob. 34.

Hence the Möbius transformations form a transformation group.

Chapter 5.2

Polynomials Over Fields

§5.2.1 Irreducible Polynomials

Df. 5.2.1.1 If for every $f(x), g(x) \in F[x]$, where $F[x]$ is an integral domain of polynomials in an indeterminate x over a field F , there exists $h(x) \in F[x]$ such that $f(x) = g(x)h(x)$, then $g(x)$ is said to be a *divisor* (or *factor*) of $f(x)$. Conversely, if $g(x)$ is a divisor of $f(x)$, then $f(x)$ is said to be *divisible* by (or a *multiple* of) $g(x)$, denoted by $g(x) \mid f(x)$ (cf. Df. 4.1.2.3.9).

Example:

$f(x) = x^5 - 1$ is a polynomial in x over a field, say, the complex number field C , and $f(x) = x^5 - 1 = (x - 1)(x^4 + x^3 + x^2 + x + 1) = g(x)h(x)$, where $f(x)$ is evidently divisible by (or a multiple of) $g(x) = x - 1$; conversely, $g(x)$ is a divisor (or factor) of $f(x)$.

This definition is merely an outcome of extending Th. 4.1.2.5.15; namely, $D[x]$ (or $I[x]$) there is extended to $F[x]$ here, which is more evident in the following *Division Algorithm*.

Th. 5.2.1.2 Given $f(x), g(x) \in F[x]$, where $g(x) \neq 0$, there uniquely exist $q(x), r(x) \in F[x]$, called the *quotient* and *remainder* of $f(x)$ divided by $g(x)$ respectively, such that

$$f(x) = q(x)g(x) + r(x), \quad 0 \leq \deg r(x) < \deg g(x)$$

where $\deg r(x)$ and $\deg g(x)$ denote the degrees of $r(x)$ and $g(x)$ respectively (cf. Df. 4.1.2.5.11). (Cf. Prob. 7.)

Example:

$3x^4 - 5x^3 + 3x - 2 = (x - 2)(3x^3 + x^2 + 2x + 7) + 12$, where $g(x) = (x - 1)$, $q(x) = 3x^3 + x^2 + 2x + 7$, and $r(x) = 12x^0$; manifestly, $\deg r(x) = 0 < \deg g(x) = 1$.

This theorem, like Th. 4.1.2.5.15, has the following immediate result:

Th. 5.2.1.3 (Remainder Theorem). For every $f(x) \in F[x]$ and $r \in R$,

$$f(x) = (x - r)g(x) + f(r)$$

where $g(x) \in F[x]$. (Cf. Prob. 8.)

Example:

$f(x) = 3x^4 - 5x^3 + 3x - 2$ implies, by Th. 5.2.1.2, $f(2) = f(x) - (x - 2)g(x) = (3x^4 - 5x^3 + 3x - 2) - (x - 2)(3x^3 + x^2 + 2x + 7) = 12$, i.e. the remainder of $f(x)$ divided by $(x - 2)$ is obtained by Th. 5.2.1.3, viz. $f(2) = 3(2)^4 - 5(2)^3 + 3(2) - 2 = 12$.

This theorem makes the following self-evident:

Th. 5.2.1.4 (Factor Theorem). For every $f(x) \in F[x]$ and $r \in F$, $f(x)$ is divisible by $x - r$, i.e. $f(x)$ has a root r , iff $f(r) = 0$. (Cf. Prob. 8 note.)

Also, parallel to Th. 4.1.2.5.18, there follows at once:

Th. 5.2.1.5 If $f(x) \in F[x]$ and $\deg f(x) = n$, then $f(x)$ has at most n distinct roots in F . (Cf. Prob. 14.)

This theorem leads to the unique factorization theorem (cf. Th. 4.1.2.3.17) in F through the following definitions and theorems.

Df. 5.2.1.6 A monic (cf. Df. 4.1.2.5.14) polynomial $d(x) \in F[x]$ is the g.c.d. (i.e. greatest common divisor) of $f(x) \neq 0$ and $g(x) \neq 0$, where $f(x), g(x) \in F[x]$, if $d(x) \mid f(x)$ and $d(x) \mid g(x)$, and if $c(x) \mid f(x)$ and $c(x) \mid g(x)$ for every $c(x) \in F[x]$ imply $c(x) \mid d(x)$. (Cf. Prob. 15.)

Th. 5.2.1.7 If $d(x)$ is the g.c.d. of $f(x)$ and $g(x)$, defined as above, then there exist $a(x), b(x) \in F[x]$ such that

$$d(x) = a(x)f(x) + b(x)g(x)$$

where $d(x)$ is the monic polynomial of least degree in this form of linear combination (cf. Th. 4.1.2.3.17). (Cf. Prob. 18-21.)

Df. 5.2.1.8 If the g.c.d. of $f(x), g(x) \in F[x]$ is 1, they are then said to be *relatively prime* (cf. Df. 4.1.2.3.15), and any polynomial $p(x) \in F[x]$ of positive degree is called *prime* (or *irreducible*) over F if it cannot be expressed as a product of two polynomials of positive degree over F .

Example:

$x^2 + 1$ and $x^2 - 2x - 3$ are relatively prime (cf. Prob. 20), and $x^2 + 1, x^3 - 5$, etc. are (individually) prime, i.e. irreducible, over F .

Th. 5.2.1.9 (Unique Factorization Theorem). Any polynomial $f(x) \in F[x]$, $\deg f(x) > 0$, can be expressed as a product:

$$a(p_1(x))^{n_1} (p_2(x))^{n_2} \cdots (p_u(x))^{n_u} \quad (D)$$

where a is the leading coefficient of $f(x)$ and $p_k(x)^{n_k}$, $k = 1, 2, \dots, u$, are monic prime polynomials over F . The expression (D) is unique except for the order of the factors. (Cf. Prob. 23.)

This theorem may be readily generalized with respect to several indeterminates; viz. the unique factorization holds in any domain $D[x_1, x_2, \dots, x_n]$ of polynomials in x_1, x_2, \dots, x_n over a domain D if the theorem holds in D . In this context D is called a *unique factorization domain* (or a *Gaussian domain*).

Df. 5.2.1.10 A polynomial with integral coefficients is called *primitive* if the g.c.d. of the coefficients is 1.

Example:

$x^2 - 1$ is primitive while $4x - 8$ is not. In general, the product of any two primitive polynomials is again primitive (cf. Prob. 25). Also:

Th. 5.2.1.11 (by Gauss). If a polynomial with integral coefficients is reducible in the rational number field R , then it can be expressed as the product of two polynomials with integral coefficients. (Cf. Prob. 26.)

On the strength of this theorem, there follows a theorem, which is the simplest of some irreducibility criteria depending on the divisibility properties of the coefficients, viz.:

Th. 5.2.1.12 (by Eisenstein). If $f(x) = \sum a_k x^k$, $k = 0, 1, \dots, t$, where $a_k \in I$ such that $a_t \not\equiv 0 \pmod{p}$, p a prime, and $a_{t-1} \equiv a_{t-2} \equiv \cdots \equiv a_0 \pmod{p}$, $a_0 \not\equiv 0 \pmod{p^2}$, then $f(x)$ is irreducible over R . (Cf. Prob. 27.)

Stated otherwise: a polynomial $f(x) = \sum a_k x^k$, $k = 0, 1, \dots, t$, is irreducible in R if there exists a prime p which is a divisor of all the a 's except a_t , and if p^2 is not a divisor of a_0 .

It goes beyond the present context, however, to find more general methods for determining the irreducibility over R of polynomials over R in general.

The concept of irreducibility or reducibility may be readily extended to the polynomials over the real number field \bar{R} or the complex number field C , as in the following theorems:

Th. 5.2.1.13 If $f(x)$, where $\deg f(x) > 2$, is a polynomial with real coefficients, then it is reducible over \bar{R} . (Cf. Prob. 33.)

Th. 5.2.1.14 If $f(x) \neq 0$ is a polynomial with real coefficients, it is uniquely factorable into $a f_1(x) f_2(x) \cdots f_n(x)$, where a is a real constant and each of the $f_k(x)$, $k = 1, 2, \dots, n$, which has real coefficients with leading coefficient 1 and of degree 1 or 2, is irreducible over \bar{R} . (Cf. Prob. 34.)

Th. 5.2.1.15 Quadratic polynomials over C are reducible over C . (Cf. Prob. 36.)

Th. 5.2.1.16 The only prime polynomials over C are the polynomials of the first degree. (Cf. Prob. 37 and Th. 5.2.3.10.)

The unique factorization theorem, together with the concept of irreducibility, is also an underlying principle in the theorems of partial fractions which are indispensable, for instance, in integral calculus.

Df. 5.2.1.17 Every element of the quotient field $Q[x]$ of polynomials in x of the integral domain $F[x]$ is of the form: $a(x)/b(x)$, which is called a *rational form* over F .

Example:

$(2x+1)/4(x^2+1)$ is evidently a rational form over F , where $(2x+1), 4(x^2+1) \in F[x]$.

It is not the case, however, that any set of rational forms constitutes a quotient field of polynomials, since, to be a field at all, it must satisfy the following theorem.

Th. 5.2.1.18 A set $F[x]/\{f(x)\}$, i.e. $F[x]$ modulo $f(x)$, of rational forms, where $F[x]$ is an integral domain of polynomials over a field F , is a field itself iff $f(x)$ is a prime polynomial over F . (Cf. Prob. 38.)

It is presumed, of course, that the following two binary operations are well-defined here for every $a(x), b(x) \in F[x]$ and some $f(x) \in F[x]$,

- (i) $\{a(x) + b(x)\} = \{a(x)\} + \{b(x)\} \pmod{f(x)}$
- (ii) $\{a(x) \cdot b(x)\} = \{a(x)\} \cdot \{b(x)\} \pmod{f(x)}$

which are patterned after the operative rules for residue classes in general (cf. Df. 4.1.2.3.18 and Th. 4.1.2.3.19-21). $F[x]/\{f(x)\}$ as such obviously contains a subfield isomorphic to the field F (cf. Prob. 39), in particular to C , expressed in the following theorem:

Th. 5.2.1.19 The quotient field $\bar{R}[x]/(x^2+1)$ is isomorphic to the complex number field C . (Cf. Prob. 40.)

Parallel to Th. 5.1.2.7, the quotient field $Q[x]$ in general yields the following theorem:

Th. 5.2.1.20 If $f(x), g(x) \in F[x]$ and $(f(x), g(x)) = 1$, then there exists $a(x) \in F[x]$ such that

$$a(x)/(f(x)g(x)) = (a(x)b(x))/f(x) + (a(x)c(x))/g(x)$$

where $b(x), c(x) \in F[x]$. (Cf. Prob. 42.)

Stated otherwise: a rational form over F , the denominator of which is the product of relatively prime polynomials $f(x)$ and $g(x)$ over F , can be expressed as a sum of two quotients, the denominators of which are $f(x)$ and $g(x)$ respectively. The rational form is then said to be expressed as a sum of *partial fractions*. This leads to:

Th. 5.2.1.21 A rational form $a(x)/b(x)$ over F can be expressed as a sum of partial fractions of the form: $q(x)/(p(x))^n$, where $p(x)$ is a prime polynomial over F , $(p(x))^n$ is a divisor of $b(x)$, and $\deg q(x) < \deg p(x)$. (Cf. Prob. 44.)

Example:

$$(x^5 + 4x^3 + x^2 + 5x - 3)/(x^4 - 1) = x/x^0 + 2/(x-1) + 3/(x+1) + (x-2)/(x^2-1). \quad (\text{Cf. Prob. 45.})$$

The prime polynomials $p(x)$ over F in this theorem must be the prime polynomials of the first or second degree if F is to be regarded as \bar{K} in particular, since this is what Th. 5.2.1.13 dictates. This conclusion is a key to the proof in the Calculus that every rational function is integrable in terms of elementary functions.

Solved Problems

1. If $f(x) \in F[x]$, then $f(x)$ is divisible by itself (i.e. a multiple of itself).

PROOF:

Since $1 = x^0 = g(x) \in F[x]$ such that

$$f(x) = f(x) \cdot 1$$

the theorem follows directly from Df. 5.2.1.1.

2. If $f(x), g(x) \in F[x]$ such that $f(x) \mid g(x)$ and $g(x) \mid f(x)$, then

$$f(x) = cg(x), \quad \deg c = 0,$$

where $c \in F$ and $c \neq 0$.

PROOF:

By hypothesis and Df. 5.2.1.1, there exist $q_1(x)$ and $q_2(x)$ in $F[x]$ such that

$$f(x) = q_1(x)g(x) \tag{1}$$

and

$$g(x) = q_2(x)f(x) \tag{2}$$

Substitute (2) in (1), and

$$\begin{aligned} f(x) &= q_1(x)(q_2(x)f(x)) = q_1(x)q_2(x)f(x) \\ \text{i.e.} \quad f(x)(1 - q_1(x)q_2(x)) &= 0 \end{aligned} \tag{3}$$

Since $F[x]$ is an integral domain (cf. Df. 5.1.2.1) which by definition cannot contain zero-divisors, it follows from (3) that

$$f(x) = 0 \tag{4}$$

or

$$q_1(x)q_2(x) = 1 \tag{5}$$

If (1) holds, then $g(x) = 0$ and the theorem is trivially true. If (2) holds, then by Th. 4.1.2.5.12,

$$\deg q_1(x) = \deg q_2(x) = 0$$

i.e. $q_1(x)q_2(x) = c$, where $\deg c = 0$, completing the proof.

3. If $c \in F$, then $cx^0 \mid f(x)$ for every $f(x) \in F[x]$.

PROOF:

If $f(x) = \sum_k a_k x^k$, $k = 0, 1, \dots, n$, then

$$f(x) = c \sum_k (a_k/c) x^k = cg(x)$$

where $g(x) = \sum_k (a_k/c) x^k$. Hence, by Prob. 2,

$$c = cx^0 \mid f(x)$$

4. If $f(x), g(x), h(x) \in F[x]$ such that $g(x) \mid f(x)$ and $h(x) \mid g(x)$, then $h(x) \mid f(x)$.

PROOF:

By hypothesis and Df. 5.2.1.1, there exist $q_1(x)$ and $q_2(x)$ in $F[x]$ such that

$$f(x) = q_1(x) g(x) \tag{1}$$

and

$$g(x) = q_2(x) h(x) \tag{2}$$

Substituting (2) in (1),

$$f(x) = q_1(x)(q_2(x) h(x)) = q_3(x) h(x)$$

where $q_3(x) \in F[x]$. Hence $h(x) \mid f(x)$.

5. If $f_1(x), f_2(x), g(x) \in F[x]$ such that $g(x) \mid f_1(x)$ and $g(x) \mid f_2(x)$, then $g(x) \mid (f_1(x) \pm f_2(x))$, and in general, if $f_1(x), f_2(x), \dots, f_n(x), g(x) \in F[x]$ such that $g(x) \mid f_1(x), g(x) \mid f_2(x), \dots, g(x) \mid f_n(x)$, then $g(x) \mid (f_1(x) \pm f_2(x) \pm \dots \pm f_n(x))$.

PROOF:

By hypothesis and Df. 5.2.1.1, there exist $q_1(x)$ and $q_2(x)$ in $F[x]$ such that

$$f_1(x) = q_1(x) g(x) \quad \text{and} \quad f_2(x) = q_2(x) g(x) \tag{1}$$

which immediately implies

$$f_1(x) \pm f_2(x) = g(x)(q_1(x) \pm q_2(x)) = q_3(x) g(x)$$

where $q_1(x) \pm q_2(x) = q_3(x) \in F[x]$. Hence $g(x) \mid (f_1(x) \pm f_2(x))$.

In general, this result justifies the assumption:

$$g(x) \mid (f_1(x) \pm f_2(x) \pm \dots \pm f_k(x))$$

which implies, by induction,

$$g(x) \mid ((f_1(x) \pm f_2(x) \pm \dots \pm f_k(x)) \pm f_{k+1}(x))$$

which justifies the final conclusion:

$$g(x) \mid (f_1(x) \pm f_2(x) \pm \dots \pm f_n(x))$$

which completes the proof.

Note. The final result may be further generalized, by Prob. 2 above, as follows: $g(x) \mid f_1(x)$, $g(x) \mid f_2(x), \dots$, and $g(x) \mid f_n(x)$ imply $g(x) \mid (\sum_i c_i f_i(x))$, $i = 1, 2, \dots, n$, where $c_i \in F$, $c_i \neq 0$, and $f_i(x) \in F[x]$.

6. If $f_1(x), f_2(x), g(x) \in F[x]$ such that $g(x) \mid f_1(x)$, then $g(x) \mid f_1(x) f_2(x)$, and in general, if $f_1(x), f_2(x), \dots, f_n(x), g(x) \in F[x]$ such that $g(x) \mid f_1(x)$, then $g(x) \mid f_1(x) f_2(x) \cdots f_n(x)$.

PROOF:

By hypothesis and Df. 5.2.1.1, there exists $q(x) \in F[x]$ such that $f_1(x) = q(x) g(x)$, which implies $f_1(x) f_2(x) = q(x) g(x) f_2(x)$, i.e.

$$g(x) \mid f_1(x) f_2(x)$$

Likewise $f_1(x) = q(x) g(x)$ implies $f_1(x) f_2(x) \cdots f_n(x) = q(x) g(x) f_2(x) \cdots f_n(x)$, i.e.

$$g(x) \mid f_1(x) f_2(x) \cdots f_n(x)$$

which completes the proof.

7. Prove Th. 5.2.1.2.

PROOF:

Let $f(x) = \sum_i a_i x^i$, $i = 0, 1, \dots, n$, and $g(x) = \sum_j b_j x^j$, $j = 0, 1, \dots, m$. Then, if $m \leq n$, multiply $g(x)$ by $a_n x^{n-m}/b_m$ and subtract it from $f(x)$ itself, i.e.

$$f(x) - (a_n x^{n-m}/b_m) g(x) = f_1(x) \quad (1)$$

where the term of the highest degree in $f(x)$, viz. $a_n x^n$, has now disappeared, i.e.,

$$f_1(x) = a_{n_1} x^{n_1} + a_{n_1-1} x^{n_1-1} + \dots + a_{0_1}$$

where $a_{n_1} \neq 0$ and $n_1 < n$. If the degree of $f_1(x)$ is still greater than or is equal to that of $g(x)$, then take a similar step to reduce the degree, viz.,

$$f_1(x) - (a_{n_1} x^{n_1-m}/b_m) g(x) = f_2(x) \quad (2)$$

and if the degree of $f_2(x)$ is still greater than or is equal to that of $g(x)$, a finite number of similar steps may be taken such that

$$f_k(x) - (a_{n_k} x^{n_k-m}/b_m) g(x) = r(x), \quad k \geq 2 \quad (3)$$

where either $r(x) = 0$ or $0 \leq \deg r(x) \leq \deg g(x)$.

Substitute (1) in (2), and take the similar steps, viz. generally substituting $(k-1)$ -th equation in (3), and it follows that

$$f(x) - q(x) g(x) = r(x)$$

where $q(x) = a_n x^{n-m}/b_m + a_{n_1} x^{n_1-m}/b_m + \dots + a_{n_k} x^{n_k-m}/b_m$. Hence

$$f(x) = q(x) g(x) + r(x), \quad 0 \leq \deg r(x) \leq \deg g(x)$$

which completes the first part of the proof.

As for the uniqueness of $q(x)$ and $r(x)$, its denial must face a contradiction. For, if there exist $q_1(x)$ and $r_1(x)$ such that

$$f(x) = q_1(x) g(x) + r_1(x), \quad 0 \leq \deg r_1(x) \leq \deg g(x)$$

then

$$q(x) g(x) + r(x) = q_1(x) g(x) + r_1(x)$$

i.e.

$$g(x)(q(x) - q_1(x)) = r_1(x) - r(x)$$

where $q(x) \neq q_1(x)$ implies $r(x) \neq r_1(x)$, which in turn implies $\deg(g(x)(q(x) - q_1(x))) > m$ while $\deg(r_1(x) - r(x)) \leq m$, contradictory to the initial assumptions.

Hence it must be the case that $q_1(x) = q(x)$, which implies $r_1(x) = r(x)$, establishing the uniqueness of $q(x)$ and $r(x)$.

8. Prove Th. 5.2.1.3.

PROOF:

Since $\deg g(x) = \deg(x-r) = 1$, the remainder $r(x)$ in Th. 5.2.1.2 must be of the degree less than 1, i.e. $\deg r(x) = 0$. Thus $r(x)$ in this context may be denoted by a constant c , $c \in F$, and

$$f(x) = (x-r) q(x) + c \quad (1)$$

Substitute $x = r$ in (1), and

$$f(r) = (r-r) q(r) + c, \quad \text{i.e.} \quad c = f(r)$$

in terms of which (1) is changed to:

$$f(x) = (x-r) g(x) + f(r) \quad (2)$$

completing the proof.

Note. From (2) above it follows at once that

$$f(x) = (x-r) g(x), \quad \text{i.e.} \quad (x-r) \mid f(x)$$

if $f(r) = 0$, and conversely, establishing Th. 5.2.1.4.

9. Divide $f(x)$ by $(x-a)(x-b)$, and find the remainder.

Solution:

(i) If $a = b$, then $f(x)$ is to be divided by $(x-a)$ twice. Hence, by this hypothesis,

$$f(x) = (x-a)g(x) + r \quad (1)$$

and

$$g(x) = (x-a)g'(x) + r' \quad (2)$$

Substitute (2) in (1), and

$$f(x) = (x-a)^2 g'(x) + r'(x-a) + r \quad (3)$$

Also, by Th. 5.2.1.3, it follows from (1) and (2) that

$$r = f(a) \quad (4)$$

which implies, by (2),

$$g(x) = (f(x) - f(a))/(x-a) \quad (5)$$

and also, from (2),

$$r' = g(a)$$

Hence the remainder is

$$r'(x-a) + r = (x-a)g(a) + f(a) \quad (6)$$

where $g(a)$ is computable by (5).

(ii) If $a \neq b$, then, by Th. 5.2.1.3,

$$f(x) = (x-a)(x-b)q(x) + r(x) \quad (7)$$

where $r(x)$ may be replaced by $cx + d$, since $g(x)$ in this context is of the second degree, viz.,

$$f(x) = (x-a)(x-b)q(x) + (cx + d) \quad (8)$$

Since (8) is an identity, substitute a , then b , in (8), and

$$f(a) = ca + d \quad \text{and} \quad f(b) = cb + d \quad (9)$$

from which c and d are readily eliminated, i.e. by solving (9) simultaneously for c and d ; thus

$$a = (f(a) - f(b))/(a-b) \quad \text{and} \quad b = (af(b) - bf(a))/(a-b)$$

Hence the remainder in this case is

$$r(x) = (f(a) - f(b))x/(a-b) + (af(b) - bf(a))/(a-b)$$

10. The remainder of the division of $f(x^2) + xg(x^2)$ by $x^2 - a$, where $f(x), g(x) \in F[x]$, is $f(a) + xg(a)$.

PROOF:

Let the quotients in the divisions of $f(t)$ and $g(t)$ by $t-a$ be $q_1(t)$ and $q_2(t)$ respectively; then, by Th. 5.2.1.3,

$$f(t) = (t-a)q_1(t) + f(a) \quad \text{and} \quad g(t) = (t-a)q_2(t) + g(a) \quad (1)$$

from which it follows, replacing t by x^2 in (1), that

$$f(x^2) = (x^2-a)q_1(x^2) + f(a), \quad g(x^2) = (x^2-a)q_2(x^2) + g(a) \quad (2)$$

from which it follows, by adding $f(x^2)$ and $xg(x^2)$, that

$$f(x^2) + xg(x^2) = (x^2-a)(q_1(x^2) + xq_2(x^2)) + f(a) + xg(a) \quad (3)$$

which proves the remainder to be $f(a) + xg(a)$, completing the proof.

11. Prove that $(x-1)^2 \mid f(x)$ if $f(x) = nx^{n+1} - (n+1)x^n + 1$.

PROOF:

By hypothesis,

$$\begin{aligned} f(x) &= nx^{n+1} - (n+1)x^n + 1 = nx^n(x-1) - (x^n-1) \\ &= (x-1)(nx^n - (x^{n-1} + x^{n-2} + \cdots + 1)) = (x-1)g(x) \end{aligned}$$

where $(x-1) \mid g(x)$, since $g(1) = n(1) - (1+1+\cdots+1) = 0$. Hence $(x-1)^2 \mid f(x)$.

12. If $g(x) \mid f(x)$, where $f(x) = ax^3 + 3bx^2 + 3cx + d$ and $g(x) = ax^2 + 2bx + c$, then $f(x)$ and $g(x)$ are of a perfect cube and a perfect square respectively.

PROOF:

If, by hypothesis, $g(x) \mid f(x)$, then, by Th. 5.2.1.4,

$$ax^3 + 3bx^2 + 3cx + d = (ax^2 + 2bx + c)(x + k)$$

where $k \in F$, must be an identity. Hence, comparing the corresponding coefficients of x^3, x^2, x^1, x^0 ,

$$3b = ak + 2b, \quad 3c = 2bk + c, \quad d = ck$$

i.e. $b = ak$, $c = bk = ak^2$, $d = ak^3$, which yield, by substitution in the original equations,

$$\begin{aligned} f(x) &= a(x^3 + 3kx^2 + 3k^2x + k^3) = a(x + k)^3 = a(x + (b/a))^3 \\ g(x) &= a(x^2 + 2kx + k^2) = a(x + k)^2 = a(x + (b/a))^2 \end{aligned}$$

completing the proof.

13. If $f(x) = \sum_k a_k x^k = 0$, $k=0, 1, \dots, n$, has n distinct roots r_1, r_2, \dots, r_n , then $f(x)$ can be expressed in the factored form $f(x) = a_n(x - r_1)(x - r_2) \cdots (x - r_n)$.

PROOF:

It follows directly from Th. 5.2.1.4 that if r_1 is a root of $f(x) = 0$, then $(x - r_1) \mid f(x)$ such that

$$f(x) = (x - r_1)f_1(x) \quad (1)$$

where $f_1(x)$ is of degree $(n-1)$ and of the form $a_n x^{n-1} + b_{n-2} x^{n-2} + \cdots + b_0$.

Conversely, if $f(x)$ has the factor $x - r_1$, then r_1 is evidently a root of $f(x)$.

Furthermore, if r_2 is also a root of $f(x)$ and $r_1 \neq r_2$ such that

$$f(r_2) = 0 \quad (2)$$

then, substituting (2) in (1),

$$(r_2 - r_1)f_1(r_2) = 0 \quad (3)$$

which implies $f_1(r_2) = 0$ ($\because r_2 - r_1 \neq 0$), which in turn implies that $(x - r_2) \mid f_1(x)$, viz.,

$$f_1(x) = (x - r_2)f_2(x) \quad (4)$$

where $f_2(x)$ is of degree $(n-2)$ and of the form $a_n x^{n-2} + c_{n-3} x^{n-3} + \cdots + c_0$. Substitute (4) in (1), and

$$f(x) = (x - r_1)(x - r_2)f_2(x)$$

verifying that $(x - r_1)(x - r_2) \mid f(x)$, where r_1 and r_2 are two distinct roots of $f(x)$.

Hence, repeating the same process,

$$f(x) = a_n(x - r_1)(x - r_2) \cdots (x - r_n) \quad (5)$$

if $f(x)$ has n distinct roots r_1, r_2, \dots, r_n .

Note. As a matter of fact, the factorization of (5) is unique, as will be verified by Th. 5.2.1.9 (cf. Prob. 23 below).

14. Prove Th. 5.2.1.5.

PROOF:

Since $f(x)$ is of degree n , by hypothesis, it follows at once from Prob. 12 that if it has n distinct roots r_1, r_2, \dots, r_n , then

$$f(x) = a_n(x - r_1)(x - r_2) \cdots (x - r_n) \quad (1)$$

Suppose, however, that $f(x)$ has $n+1$ distinct roots, say r_{n+1} in addition to the original n roots r_1, r_2, \dots, r_n . Then, substituting $x = r_{n+1}$ in (1), it follows that, by Th. 5.2.1.4,

$$f(r_{n+1}) = a_n(r_{n+1} - r_1)(r_{n+1} - r_2) \cdots (r_{n+1} - r_n) = 0$$

which is contradictory to the assumption that $r_{n+1} \neq r_1, r_{n+1} \neq r_2, \dots, r_{n+1} \neq r_n$.

Hence $f(x)$ cannot have more than n distinct roots, which completes the proof.

15. Prove the *Euclidean Algorithm* (cf. §4.1.2.3, Prob. 31) with respect to $F[x]$, viz. $f(x)$ and $g(x)$, where $f(x), g(x) \in F[x]$ and $f(x) \neq 0, g(x) \neq 0$, have a g.c.d. $d(x) \in F[x]$.

PROOF:

Let $\deg f(x) \geq \deg g(x)$; then, by Th. 5.2.1.3,

$$f(x) = q(x)g(x) + r(x), \quad 0 \leq \deg r(x) \leq \deg g(x) \quad (1)$$

If $r(x) = 0$, then a g.c.d. of $f(x)$ and $g(x)$ is $g(x)$ itself.

If $r(x) \neq 0$, then apply the division algorithm of Th. 5.2.1.3 to $g(x)$ and $r(x)$, and

$$g(x) = q_1(x)r(x) + r_1(x), \quad 0 \leq \deg r_1(x) \leq \deg r(x) \quad (2)$$

If $r_1(x) = 0$, then $r(x)$ itself is a g.c.d. of $g(x)$ and $r(x)$, hence a g.c.d. of $f(x)$ and $g(x)$.

If $r_1(x) \neq 0$, then the algorithm must be repeated, viz.,

$$r(x) = q_2(x)r_1(x) + r_2(x), \quad 0 \leq \deg r_2(x) \leq \deg r_1(x) \quad (3)$$

and in general,

$$r_k(x) = q_{k+2}(x)r_{k+1}(x) + r_{k+2}(x), \quad 0 \leq \deg r_{k+2}(x) \leq \deg r_{k+1}(x) \quad (4)$$

until it reaches the last stages,

$$r_{n-2}(x) = q_n(x)r_{n-1}(x) + r_n(x), \quad 0 \leq \deg r(x) \leq \deg r(x) \quad (5)$$

and

$$r_{n-1}(x) = q_{n+1}(x)r_n(x) \quad (6)$$

where the remainder is finally zero, in which case $q_{n+1}(x)$ may be at least a polynomial of degree 0, say $ax^0 \in F$, which certainly divides every $f(x)$ for $f(x) \in F[x]$. Hence, by (1)-(6), it follows that

$$\begin{aligned} (f(x), g(x)) &= (g(x), r(x)) = (r(x), r_1(x)) = \cdots \\ &= (r_k(x), r_{k+1}(x)) = \cdots \\ &= (r_{n-2}(x), r_{n-1}(x)) = (r_{n-1}(x), r_n(x)) = r_n(x) \end{aligned}$$

i.e. $r_n(x) = d(x)$ is a g.c.d. of $f(x)$ and $g(x)$.

Note. Since $(f(x), g(x)) = d_1(x)$ and $(g(x), r(x)) = d_2(x)$ imply $d_2(x) \mid d_1(x)$ and $d_1(x) \mid d_2(x)$ in the context of (1) and (2) above, $d_1(x)$ and $d_2(x)$ are *associates*, by Df. 4.1.2.3.12, and differ only by a factor, say, $c \in F$.

16. Find a g.c.d. of $x^3 + x^2 - 2$ and $x^3 + 2x^2 - 3$.

Solution:

$$\begin{array}{r} x^3 + x^2 - 2 \mid x^3 + 2x^2 - 3 \quad \underline{1} \\ \underline{x^3 + x^2 - 2} \\ x^2 - 1 \mid x^3 + x^2 - 2 \quad \underline{x+1} \\ \underline{x^3 - x} \\ x^2 + x - 2 \\ \underline{x^2 - 1} \\ x - 1 \mid x^2 - 1 \quad \underline{x+1} \\ \underline{x^2 - x} \\ x - 1 \\ \underline{x - 1} \\ 0 \end{array}$$

Hence $((x^3 + x^2 - 2), (x^3 + 2x^2 - 3)) = x - 1$.

Note. As is quite obvious, the algorithm may be carried out in the form of synthetic division, as in Prob. 17 below.

17. Find a g.c.d. of $x^3 - x^2 - x + 1$ and $x^4 - 3x^2 - 2x + 4$.

Solution:

By Prob. 15,

$$\begin{array}{r|l}
 \begin{array}{r}
 g(x) \quad \begin{array}{rrrr} 1 & -1 & -1 & 1 \end{array} \\
 \begin{array}{rrrr} 1 & 2 & -3 & \end{array} \\
 \hline
 \begin{array}{rrrr} -3 & 2 & 1 & \end{array} \\
 \begin{array}{rrrr} -3 & -6 & 9 & \end{array} \\
 \hline
 \begin{array}{rrrr} & 8 & -8 & \end{array} \\
 \div 8) \hline
 \begin{array}{rrrr} & 1 & -1 & \end{array}
 \end{array}
 &
 \begin{array}{r}
 \begin{array}{rrrrr} 1 & 0 & -3 & -2 & 4 \end{array} \\
 \begin{array}{rrrrr} 1 & -1 & -1 & 1 & \end{array} \\
 \hline
 \begin{array}{rrrrr} & 1 & -2 & -3 & 4 \end{array} \\
 \begin{array}{rrrrr} & 1 & -1 & -1 & 1 \end{array} \\
 \hline
 \begin{array}{rrrrr} & & -1 & -2 & 3 \end{array} \\
 \begin{array}{rrrrr} & & -1 & 1 & \end{array} \\
 \hline
 \begin{array}{rrrrr} & & & -3 & 3 \end{array} \\
 \begin{array}{rrrrr} & & & -3 & 3 \end{array} \\
 \hline
 \begin{array}{rrrrr} & & & & 0 \end{array}
 \end{array}
 \end{array}
 \begin{array}{l}
 f(x) \\
 \\
 \\
 \\
 \\
 \\
 r(x)
 \end{array}$$

Hence $((x^3 - x^2 - x + 1), (x^4 - 3x^2 - 2x + 4)) = x - 1$.

18. Prove Th. 5.2.1.7.

PROOF:

By hypothesis and Th. 5.2.1.3,

$$f(x) = q(x)g(x) + r(x)$$

i.e.

$$r(x) = f(x) - q(x)g(x) \quad (1)$$

from which, and by Prob. 15, (2), it follows that

$$r_1(x) = g(x) - q_1(x)r(x) = g(x) - q_1(x)(f(x) - q(x)g(x)) = (-q_1(x))f(x) + (1 + q(x)q_1(x))g(x)$$

Assume, then, that the equation

$$r_n(x) = s_n(x)f(x) + t_n(x)g(x) \quad (2)$$

is valid up to the i th repetitive process, viz.,

$$r_i(x) = s_i(x)f(x) + t_i(x)g(x), \quad i = 0, 1, \dots, i < k \quad (3)$$

where (1) above is evidently the case of (3) for $i = 0$, letting $r(x) = r_0(x)$, $s_0(x) = 1$, $t_0(x) = -q(x)$; (2) may be considered likewise. Hence, in general, by (3) above and Prob. 15, (4),

$$r_{k-2}(x) = r_{k-1}(x)q_k(x) + r_k(x) \quad (4)$$

and solving (4) for $r_k(x)$ and substituting the values for $r_{k-1}(x)$ and $r_{k-2}(x)$ as given by the induction hypothesis of (3), it follows that

$$\begin{aligned}
 r_k(x) &= -q_k(x)(s_{k-1}(x)f(x) + t_{k-1}(x)g(x)) + (s_{k-2}(x)f(x) + t_{k-2}(x)g(x)) \\
 &= (-q_k(x)s_{k-1}(x) + s_{k-2}(x))f(x) + (-q_k(x)t_{k-1}(x) + t_{k-2}(x))g(x)
 \end{aligned}$$

where $(-q_k(x)s_{k-1}(x) + s_{k-2}(x))$ and $(-q_k(x)t_{k-1}(x) + t_{k-2}(x))$ may be replaced by $a(x)$ and $b(x)$ respectively, and in the same general term, $r_k(x)$ may be replaced by $d(x)$, by hypothesis, viz.

$$d(x) = a(x)f(x) + b(x)g(x)$$

which completes the proof.

19. Given $f(x) = x^4 - 3x^2 - 2x - 4$ and $g(x) = x^3 - x^2 - x + 1$ (as in Prob. 17), find $a(x)$ and $b(x)$ such that

$$x - 1 = a(x)f(x) + b(x)g(x)$$

Solution:

Let $f(x) = f_0$ and $g(x) = f_1$; then, by Prob. 15,

$$f_0 = f_1q_1 + f_2, \quad f_1 = f_2q_2 + f_3, \quad f_2 = f_3q_3$$

where $q_1 = x + 1$, $f_2 = -x^2 - 2x + 3$, $q_2 = -x + 3$, $f_3 = 8x - 8$ (cf. Prob. 17 above), which implies

$$f_3 = f_1 - f_2q_2 = f_1 - (f_0 - f_1q_1)q_2 = -f_0q_2 + f_1(1 + q_1q_2)$$

i.e.

$$8x - 8 = (x - 3)f(x) + (1 - (x + 1)(x - 3))g(x)$$

which immediately yields the required equations.

$$a(x) = (x - 3)/8 \quad \text{and} \quad b(x) = (-x^2 + 2x + 4)/8$$

20. Given $f(x) = x^2 + 1$ and $g(x) = x^2 - 2x - 3$, find $a(x)$ and $b(x)$ such that $a(x)f(x) + b(x)g(x) = 1$.

Solution:

Since, by the Euclidean Algorithm,

$$x^2 + 1 = (x^2 - 2x - 3) \cdot 1 + (2x + 4) \quad \text{and} \quad x^2 - 2x - 3 = (2x + 4) \cdot ((x - 4)/2) + 5$$

it follows immediately that

$$\begin{aligned} 5 &= (x^2 - 2x - 3) - (2x + 4)(x - 4)/2 = (x^2 - 2x - 3) - ((x^2 + 1) - (x^2 - 2x - 3))(x - 4)/2 \\ &= -(x^2 + 1)(x - 4)/2 + (x^2 - 2x - 3)(1 + ((x - 4)/2)) \end{aligned}$$

or

$$1 = ((4 - x)/10)(x^2 + 1) + ((x - 2)/10)(x^2 - 2x - 3)$$

which implies that

$$a(x) = (4 - x)/10 \quad \text{and} \quad b(x) = (x - 2)/10$$

21. If $d(x) = (f(x), g(x))$ and if $a(x)$ and $b(x)$ such that

$$d(x) = a(x)f(x) + b(x)g(x) \tag{1}$$

are already known, then any $a'(x)$ and $b'(x)$ such that

$$d(x) = a'(x)f(x) + b'(x)g(x) \tag{2}$$

are given by the following relations:

$$a'(x) = a(x) - u(x)t(x) \quad \text{and} \quad b'(x) = b(x) + u(x)s(x) \tag{3}$$

where $s(x) = f(x)/d(x)$, $t(x) = g(x)/d(x)$, and $u(x)$ is an arbitrary polynomial in $F[x]$.

PROOF:

From (1) and (2),

$$f(x)(a(x) - a'(x)) = g(x)(b'(x) - b(x)) \tag{4}$$

Divide (4) by $d(x)$; then, by hypothesis,

$$s(x)(a(x) - a'(x)) = t(x)(b'(x) - b(x)) \tag{5}$$

Since $(s(x), t(x)) = 1$, by hypothesis, which in turn implies $s(x) \mid (t(x)(b'(x) - b(x)))$, it follows that, for some $u(x) \in F[x]$, $b'(x) - b(x) = u(x)s(x)$, and, consequently, $a(x) - a'(x) = u(x)t(x)$, which yields (3).

Conversely, if (3) holds, then

$$\begin{aligned} a'(x)f(x) + b'(x)g(x) &= f(x)(a(x) - u(x)t(x)) + g(x)(b(x) + u(x)s(x)) \\ &= a(x)f(x) + b(x)g(x) + (s(x)g(x) - t(x)f(x))u(x) \\ &= d(x) + (d(x)t(x)s(x) - d(x)s(x)t(x))u(x) \\ &= d(x) \end{aligned}$$

which completes the proof.

Note. According to this theorem, all other possible sets of $a'(x)$ and $b'(x)$ in Prob. 20 are given as follows:

$$a'(x) = (4 - x)/10 - (x^2 - 2x - 3)u(x), \quad b'(x) = (x - 2)/10 + (x^2 + 1)u(x)$$

for some $u(x) \in F[x]$.

22. If $h(x) \in F[x]$ is irreducible and divides the product $f(x)g(x)$ of $f(x)$ and $g(x)$ in $F[x]$, then $h(x) \mid f(x)$ or $h(x) \mid g(x)$.

PROOF:

Assume $h(x) \nmid f(x)$. Since $h(x)$ is irreducible, by hypothesis, there are no other divisors of $h(x)$ but its associates (cf. Prob. 15, note) or units of F , which implies $(f(x), h(x)) = 1$, which in turn implies, by Th. 5.2.1.7,

$$1 = a(x)f(x) + b(x)h(x) \tag{1}$$

for some $a(x), b(x) \in F[x]$. Multiply (1) by $g(x)$,

$$g(x) = a(x)f(x)g(x) + b(x)h(x)g(x) \quad (2)$$

which implies that, since $h(x) \mid f(x)g(x)$ by hypothesis, $h(x)$ must be a factor of both sides of (2), viz., $h(x) \mid g(x)$.

On the other hand, if $h(x) \nmid g(x)$, then a similar reasoning yields $h(x) \mid f(x)$, which completes the proof.

Note. This theorem is readily generalized to the case of an irreducible polynomial $p(x)$ with respect to n polynomials $f_1(x), f_2(x), \dots, f_n(x)$.

23. Prove Th. 5.2.1.9.

PROOF:

If $f(x)$ is irreducible, then the theorem is already complete as such.

If $f(x)$ is reducible, then let

$$f(x) = f_1(x)f_2(x) \quad (1)$$

where evidently $\deg f_1(x) < \deg f(x)$ and $\deg f_2(x) < \deg f(x)$, which implies an induction that such a decomposition as (1) may be repeated, yielding polynomials of degree less than $\deg f(x)$, viz.,

$$f_1(x) = cg_1(x)g_2(x) \cdots g_r(x) \quad \text{and} \quad f_2(x) = dh_1(x)h_2(x) \cdots h_s(x) \quad (2)$$

where $c, d \in F$, and $g_i(x)$, $i = 1, 2, \dots, r$, and $h_j(x)$, $j = 1, 2, \dots, s$, are monic irreducible polynomials over F . Hence, by (1) and (2),

$$f(x) = cdg_1(x)g_2(x) \cdots g_r(x)h_1(x)h_2(x) \cdots h_s(x) = a(p_1(x))^{n_1}(p_2(x))^{n_2} \cdots (p_u(x))^{n_u} \quad (3)$$

where $cd = a \in F$, which is the leading coefficient of (x) in decomposition, and $p_k(x)^{n_k}$, $k = 1, 2, \dots, u$, may be any of $g_i(x)$ and $h_j(x)$, some of which may be identical, thus yielding the exponents $n_k \in I$, $k = 1, 2, \dots, u$, where obviously $1 \leq n_k \leq \deg f(x)$.

Furthermore, the decomposition of (3) is unique. For, if there exist two decompositions for $f(x)$, viz.,

$$f(x) = a(p_1(x))^{n_1}(p_2(x))^{n_2} \cdots (p_u(x))^{n_u} = b(q_1(x))^{m_1}(q_2(x))^{m_2} \cdots (q_v(x))^{m_v} \quad (4)$$

then $a = b$, since these prime polynomials are monic. Also, since any of the p 's, say $p_1(x)$, is prime, i.e. irreducible, (4) implies that $p_1(x)$ must divide some of the q 's, say $q_1(x)$. But, both being monic and prime by the initial assumption, their quotient cannot but be the unity of F ; hence $p_1(x) = q_1(x)$. Dividing (4) by this common factor and a , there follows from (4):

$$(p_2(x))^{n_2}(p_3(x))^{n_3} \cdots (p_u(x))^{n_u} = (q_2(x))^{m_2}(q_3(x))^{m_3} \cdots (q_v(x))^{m_v}$$

where similar steps of elimination may be repeated such that

$$p_1(x) = q_1(x), \quad p_2(x) = q_2(x), \quad \dots, \quad p_u(x) = q_v(x) \quad (5)$$

can be established, although the order of the factors may not be the same as in (5). In any case, except for the order of the factors, the decomposition of (3) is thus unique, which completes the proof.

24. $f(x) = x^3 + y^3 + xy$ is irreducible in the complex number field C .

PROOF:

If $f(x)$ is reducible, then it must have a linear factor, viz.,

$$f(x) = x^3 + y^3 + xy = (x + ay + b)(x^2 + cxy + dy^2 + rx + sy + t) \quad (1)$$

where $a, b, c, d, r, s, t \in C$. Let $x = 0$; then $y^3 = (ay + b)(dy^2 + sy + t)$, which implies $ad = 1$ and $b = s = t = 0$; likewise $y = 0$ implies $r = 0$. Hence (1) is equivalent to

$$x^3 + y^3 + xy = (x + ay)(x^2 + cxy + dy^2) \quad (2)$$

which is a contradiction, since the right-hand side of (2) is homogeneous while the left-hand side is not. Hence, contrary to the initial assumption, $f(x)$ must be irreducible, which was to be proved.

25. The product of two primitive polynomials is again a primitive polynomial.

PROOF:

Let $f(x), g(x), h(x)$ be polynomials such that

$$f(x) = g(x)h(x) \quad (1)$$

where

$$\begin{aligned} f(x) &= \sum_k a_k x^k, \quad k = 0, 1, \dots, t, \\ g(x) &= \sum_i b_i x^i, \quad i = 0, 1, \dots, r, \\ h(x) &= \sum_j c_j x^j, \quad j = 0, 1, \dots, s, \end{aligned} \quad (2)$$

and $a_k, b_i, c_j \in I$, $a_t b_r c_s \neq 0$, such that $t = r + s$. Then, by direct multiplication, the relations among the coefficients of $f(x), g(x), h(x)$ are found to be:

$$\begin{aligned} a_t &= b_r c_s, \\ a_{t-1} &= b_{r-1} c_s + b_r c_{s-1}, \\ a_{t-2} &= b_{r-2} c_s + b_{r-1} c_{s-1} + b_r c_{s-2}, \\ &\dots\dots\dots \\ a_s &= b_0 c_s + b_1 c_{s-1} + \dots\dots\dots \\ &\dots\dots\dots \\ a_2 &= b_0 c_2 + b_1 c_1 + b_2 c_0, \\ a_1 &= b_0 c_1 + b_1 c_0, \\ a_0 &= b_0 c_0 \end{aligned} \quad (3)$$

thus in general

$$a_k = b_0 c_k + b_1 c_{k-1} + \dots + b_{k-1} c_1 + b_k c_0 \quad (4)$$

where $b_i = 0$ and $c_j = 0$ if $i > r$ and $j > s$, allowing the following alternative of (4):

$$a_{m+n} = b_m c_n + b_{m-1} c_{n+1} + \dots + b_{m+1} c_{n-1} + b_{m+2} c_{n-2} + \dots \quad (5)$$

Let, by hypothesis, $g(x)$ and $h(x)$ be primitive, i.e. $(b_r, b_{r-1}, \dots, b_0) = 1$ and $(c_s, c_{s-1}, \dots, c_0) = 1$, and assume, contrary to the desired conclusion, that $(a_t, a_{t-1}, \dots, a_0) \neq 1$. Then there exists a prime number p which is a divisor of all the a 's, implying that both $g(x)$ and $h(x)$ must have at least one coefficient that cannot be divided by p , since $g(x)$ and $h(x)$ are given as primitive polynomials. Let, then, b_m and c_n be of the smallest subscripts of $g(x)$ and $h(x)$ such that $p \nmid b_m$ and $p \nmid c_n$, where obviously $m, n \geq 1$. But, then, it follows from (5) that $p \mid b_m c_n$, hence $p \mid b_m$ or $p \mid c_n$, since $p \mid a_{m+n}$, $b_{m+1}, b_{m+2}, \dots, c_{n+1}, c_{n+2}, \dots$. This contradiction yields $(a_t, a_{t-1}, \dots, a_0) = 1$, which completes the proof.

26. Prove Th. 5.2.1.11.

PROOF:

If $f(x)$ is given, then by hypothesis,

$$f(x) = g(x)h(x) \quad (1)$$

where $g(x), h(x) \in R[x]$ and $\deg g(x), \deg h(x) \geq 1$. If, furthermore, r and s are the least common denominators of the coefficients of $g(x)$ and $h(x)$ respectively, then

$$g'(x) = r g(x) \quad \text{and} \quad h'(x) = s h(x) \quad (2)$$

which together imply

$$rs f(x) = g'(x)h'(x) \quad (3)$$

where $g'(x)$ and $h'(x)$ are polynomials with integral coefficients. Now, if each of a, b, c represents the g.c.d. of $f(x), g(x), h(x)$ respectively,

$$f(x)/a, \quad g''(x) = g'(x)/b, \quad h''(x) = h'(x)/c \quad (4)$$

are primitive, by Df. 5.2.1.10, and so, by Prob. 24, is their product:

$$g''(x)h''(x) = (g'(x)/b)(h'(x)/c) = (f(x)/a)(ars/bc)$$

by (3) and (4). Hence

$$ars/bc = 1 \quad (5)$$

which implies, by (3), (4), (5),

$$f(x) = g'(x)h'(x)/rs = b g''(x) c h''(x)/rs = (bc/rs) g''(x) h''(x) = a g''(x) h''(x)$$

where $ag''(x)$ and $h''(x)$ are evidently polynomials with integral coefficients, whose degrees are the same as those of $g(x)$ and $h(x)$ respectively. This completes the proof.

27. Prove Th. 5.2.1.12.

PROOF:

From Th. 5.2.1.11 it follows that

$$f(x) = g(x)h(x)$$

where $g(x)$ and $h(x)$ are also polynomials with integral coefficients and $\deg g(x), \deg h(x) \geq 1$.

Let $f(x), g(x), h(x)$ be represented as in (2) of Prob. 25 above. Then, since $p \mid a_0$ and $p^2 \nmid a_0$, by hypothesis, it follows that either $p \mid b_0$ or $p \mid c_0$, but not both. Assume, say, $p \nmid b_0$ and $p \mid c_0$; then, in the equation of $a_1 = b_0 c_1 + b_1 c_0$ in (3), it follows that $p \mid c_1$, since $p \mid a_1$, $p \mid c_0$, and $p \nmid b_0$. Likewise $p \mid c_2$ in $a_2 = b_0 c_2 + b_1 c_1 + b_2 c_0$, and in general

$$a_s = b_0 c_s + b_1 c_{s-1} + \dots$$

yields $p \mid c_{s-1}$, which finally, at the end of a finite number of similar steps, yields $p \mid c_s$, which at once implies $p \mid a_s$, contradicting the original assumption.

Since the conclusion is dually the same for assuming $p \mid b_0$, it must follow that $f(x)$ is irreducible in R , completing the proof.

28. If p is a prime and $f(x) = x^{p-1} + x^{p-2} + \dots + x + 1$, then $f(x)$ is irreducible in R .

PROOF:

Since, by hypothesis, $(x-1)f(x) = x^p - 1$, let $x = y+1$; then $y f(y+1) = (y+1)^p - 1$, i.e.,

$$f(y+1) = y^{p-1} + {}_p C_1 y^{p-2} + {}_p C_2 y^{p-3} + \dots + {}_p C_{p-2} y + {}_p C_{p-1} \quad (1)$$

where ${}_p C_r = p(p-1)\dots(p-r+1)/r!$, which is of course divisible by p .

Assume that $f(x)$, i.e. $f(y+1)$, is reducible in R . Then, by Prob. 26, there follows:

$$f(y+1) = g(y+1)h(y+1)$$

where $g(y+1)$ and $h(y+1)$ are polynomials with integral coefficients, which is a contradiction, since $p \mid {}_p C_r$, $p \nmid {}_p C_0$ ($\therefore {}_p C_0 = 1$) and $p^2 \nmid {}_p C_{p-1}$ ($\therefore {}_p C_{p-1} = p$) in (1). This implies, by Th. 5.2.1.12, that $f(y+1)$, i.e. $f(x)$, is irreducible in R , completing the proof.

Note. The transformation of $f(x)$ into $f(y+1)$ by $x = y+1$ above is fully justified; for in general, by Prob. 26, $a(x) = b(x)c(x)$ implies $a(x+1) = b(x+1)c(x+1)$, and $a(x+1) = b'(x)c'(x)$ implies $a(x) = b'(x-1)c'(x-1)$. Hence $a(x)$ and $a(x+1)$ are simultaneously reducible or irreducible.

29. If $a(x), b(x) \in F[x]$, where $a(x) \neq 0$ and $b(x)$ is irreducible over F , and if $a(x)$ and $b(x)$ have a root in common, then there exists $c(x) \in F[x]$ such that $a(x) = (b(x))^n c(x)$, where $n \in N$, and $b(x)$ and $c(x)$ have no root in common.

PROOF:

Let $d(x) \in F[x]$ be a highest common factor of $a(x)$ and $b(x)$ in $F[x]$; then, by this assumption, every root of $a(x)$ and $b(x)$ is a root of $d(x)$, and also $\deg d(x) \geq 1$, since $a(x)$ and $b(x)$ do have a root in common by hypothesis. Now, also by hypothesis, $b(x)$ is irreducible over F , which implies that $b(x)$ has no factors in $F[x]$ except its associates and constants. Since $\deg d(x) \geq 1$, $d(x)$ must then be an associate of $b(x)$, which implies that $b(x)$ itself must be a factor of $a(x)$. Hence $a(x) = (b(x))^n c(x)$, $n \in N$ representing the highest power of $b(x)$ which divides $a(x)$, for some $c(x) \in F(x)$.

Furthermore, since $c(x) \nmid a(x)$ in compliance with the stipulation with respect to n in the above context, $b(x)$ and $c(x)$ cannot have any root in common, which completes the proof.

30. If $f(x) \in F[x]$ has real coefficients such that $f(x) = (x - (a + ib))^n g(x)$ for some $n \in N$ and some $g(x) \in F[x]$, where $a, b \in \bar{R}$ are not simultaneously zero, then

$$f(x) = (x - (a + ib))^n (x - (a - ib))^n g(x)$$

PROOF:

Let $h(x) = (x - (a + ib))(x - (a - ib)) = x^2 - 2ax + a^2 + b^2$, where $h(x)$ has evidently only real coefficients and, having no real root, is irreducible over \bar{R} . Then, since $f(x)$ and $h(x)$ have the common root $a + ib$, it follows immediately from Prob. 29 above that, for some $n \in N$ and some $g(x) \in F[x]$,

$$f(x) = (h(x))^n g(x) = (x - (a + ib))^n (x - (a - ib))^n g(x)$$

which completes the proof.

31. If $f(x) = x^4 + 3x^3 + 6x^2 + 12x + 8$ is known to have a root $-2i$, then find all roots of $f(x)$.

Solution:

Since $-2i$ is a root of $f(x)$, i.e. $(x - (-2i)) \mid f(x)$, it follows from Prob. 30 that $(x - 2i) \mid f(x)$. Hence $(x + 2i)(x - 2i) \mid f(x)$, and by division

$$f(x) = (x^2 + 4)(x^2 + 3x + 2) = (x + 2i)(x - 2i)(x + 1)(x + 2)$$

which yields all four roots of $f(x)$: $-2i, 2i, -1, -2$.

32. Determine a and b in $f(x) = x^3 - 6x^2 + ax + b$ such that $f(x)$ has a root $1 - i\sqrt{5}$, and then solve $f(x)$.

Solution:

Since Prob. 30 dictates that $f(x)$ must have also $1 + i\sqrt{5}$ as a root, $f(x)$ must be divided by $(x - (1 - i\sqrt{5}))(x - (1 + i\sqrt{5})) = x^2 - 2x + 6 = g(x)$. Divide $f(x)$ by $g(x)$, i.e.

$$\begin{array}{r} 1 - 4 \\ 1 - 2 + 6 \overline{) 1 - 6 \quad a \quad b} \\ \underline{1 - 2 \quad 6} \\ - 4 \quad a - 6 \quad b \\ \underline{- 4 \quad 8 \quad -24} \\ a - 14 \quad b + 24 \end{array}$$

which implies $a - 14 = b + 24 = 0$, i.e. $a = 14$ and $b = -24$. Hence

$$f(x) = x^3 - 6x^2 + 14x - 24 = (x - 4)(x^2 - 2x + 6)$$

which yields 4 and $1 \pm i\sqrt{5}$ for the roots of $f(x)$.

33. Prove Th. 5.2.1.13.

PROOF:

Since the fundamental theorem of algebra (cf. Th. 5.2.3.9) assures the existence of a root, say a , for $f(x)$, it follows at once, from Th. 5.2.1.4, that $(x - a) \mid f(x)$.

If $a \in \bar{R}$, it implies *ipso facto* that $f(x)$ has a factor with real coefficients.

If $a \in C$, say $a = c + id$, then, by Prob. 30 above,

$$(x - (c + id))(x - (c - id)) \mid f(x)$$

where the factor has real coefficients.

In either case $f(x)$ has thus a factor, say $g(x)$, which has real coefficients and, of course, $\deg g(x) \leq \deg f(x)$, where $\deg f(x) > 2$. Hence $f(x)$ is reducible over \bar{R} , completing the proof.

34. Prove Th. 5.2.1.14.

PROOF:

The first part of the theorem follows directly from Th. 5.2.1.9, and the second from Th. 5.2.1.13, which, stated otherwise, concludes that a prime polynomial $p(x)$ over \bar{R} implies $\deg p(x) \leq 2$.

(Th. 5.2.1.14, then, merely yields the combined effect of both Th. 5.2.1.9 and Th. 5.2.1.13.)

35. Quadratic polynomials over \bar{R} are prime iff their discriminants are negative.

PROOF:

Let $f(x) = ax^2 + bx + c$, where $a, b, c \in \bar{R}$, which yields the discriminant of $f(x)$: $D = \sqrt{b^2 - 4ac}$; then, since

$$f(x) = a(x - r_1)(x - r_2) = a(x - ((-b + \sqrt{b^2 - 4ac})/2a))(x - ((-b - \sqrt{b^2 - 4ac})/2a))$$

it follows that $D < 0$ implies $r_1, r_2 \notin \bar{R}$. Hence $f(x)$ is irreducible over \bar{R} if $D < 0$.

Conversely, if $f(x)$ is prime over \bar{R} , it must be the case that $D < 0$, since both $D > 0$ and $D = 0$ imply $r_1, r_2 \in \bar{R}$, in which cases $f(x)$ is evidently not irreducible over \bar{R} . Hence, by the law of trichotomy, the irreducibility of $f(x)$ over \bar{R} implies $D < 0$, completing the proof.

36. Prove Th. 5.2.1.15.

PROOF:

Let the quadratic polynomials over C be represented by $f(x) = ax^2 + bx + c$, where $a, b, c \in C$; then $f(x) = a(x - r_1)(x - r_2)$, where r_1 and r_2 are two roots of $f(x)$ and are of the form

$$r_1 = (-b + \sqrt{b^2 - 4ac})/2a, \quad r_2 = (-b - \sqrt{b^2 - 4ac})/2a \quad (1)$$

which are complex numbers (cf. §5.1.3, Prob. 28-29). Thus the first-degree factors of $f(x)$ with complex coefficients do exist, and this defies any irreducibility over C of quadratic polynomials over C , which completes the proof.

Note. $r_1 = r_2$ iff $\sqrt{b^2 - 4ac} = 0$, parallel to the case where $f(x)$ is defined over \bar{R} (cf. Prob. 35 above).

37. Prove Th. 5.2.1.16.

PROOF:

The fundamental theorem of algebra (cf. Th. 5.2.3.9) implies that $f(x) \in C(x)$ of positive degree has a root $c \in C$. Hence, by Th. 5.2.1.4, $f(x) = (x - c)g(x)$, where $\deg g(x) = n - 1$ if $\deg f(x) = n$. It is evident, then, that $f(x)$ cannot be prime over C if $\deg f(x) \geq 2$, which completes the proof.

38. Prove Th. 5.2.1.18.

PROOF:

Once the binary operations for $F[x]/\{f(x)\}$ are defined, viz. as follows:

$$\{a(x) + b(x)\} = \{a(x)\} + \{b(x)\} \pmod{f(x)}$$

$$\{a(x) \cdot b(x)\} = \{a(x)\} \cdot \{b(x)\} \pmod{f(x)}$$

for every $a(x), b(x) \in F[x]$, it follows readily that

$$\{a(x)\} + \{b(x)\} = \{b(x)\} + \{a(x)\} \pmod{f(x)}$$

$$\{a(x)\} \cdot \{b(x)\} = \{b(x)\} \cdot \{a(x)\} \pmod{f(x)}$$

etc., satisfying D1-11 one after another (cf. Df. 4.1.2.2.1) with $\{0\}$ for D3 and $\{1\}$ for D10. The residue class $F[x]/\{f(x)\}$ is thus an integral domain.

Furthermore, if $f(x)$ is a prime polynomial over F , then, by Th. 5.2.1.7, there must exist some $a(x), b(x), g(x) \in F[x]$ such that

$$a(x)f(x) + b(x)g(x) = 1$$

where $g(x) \neq 0$ and $(f(x), g(x)) = 1$, which implies

$$b(x) \cdot g(x) \equiv 1 \pmod{f(x)}$$

i.e.

$$\{g(x)\} \cdot \{b(x)\} = \{1\}$$

which in turn implies the existence of a multiplicative inverse of $g(x)$.

This establishes F10 (cf. Df. 4.1.2.4.1) for the integral domain $F[x]/\{f(x)\}$, which is thus a field.

Conversely, if the residue class $F[x]/\{f(x)\}$ is a field, then $f(x)$ must be a prime polynomial over F . Or, what is the same, if $f(x)$ is not a prime polynomial over F , $F[x]/\{f(x)\}$ is not a field; in fact, it is not even an integral domain. For, if $f(x)$ is not prime over F , then $f(x) = f_1(x)f_2(x)$, where $\deg f_1(x) < \deg f(x)$ and $\deg f_2(x) < \deg f(x)$, which evidently imply $f_1(x) \nmid f(x)$ and $f_2(x) \nmid f(x)$, i.e. $\{f_1(x)\} \neq \{0\}$ and $\{f_2(x)\} \neq \{0\}$, while

$$\{f_1(x)\} \cdot \{f_2(x)\} = \{f_1(x)f_2(x)\} = \{f(x)\} = \{0\}$$

which defies D11 (cf. Df. 4.1.2.2.1). Hence the class $F[x]/\{f(x)\}$ is not an integral domain, let alone a field, if $f(x)$ is not prime over F . This completes the proof.

39. The residue class $F[x]/\{f(x)\}$, defined in Prob. 38 above, contains a subclass which is isomorphic to F .

PROOF:

Let G be the set of all elements of the form: $\{a\}$, where $a \in F$, of the residue class $F[x]/\{f(x)\}$, which has been proved to be a field (cf. Prob. 38 above). Then the mapping

$$M: \{a\} \leftrightarrow a$$

is a 1-1 mapping of G into F , since $\{a\} = \{b\}$ iff $\{a\} \equiv \{b\} \pmod{f(x)}$, which in turn holds iff $a = b$. Also, $\{a\} \leftrightarrow a$ and $\{b\} \leftrightarrow b$ imply

$$\begin{aligned} M_1: \{a\} + \{b\} \pmod{f(x)} &\leftrightarrow a + b \\ M_2: \{a\} \cdot \{b\} \pmod{f(x)} &\leftrightarrow a \cdot b \end{aligned}$$

verifying the mapping M to be an isomorphism of G , a subclass of the given residue class, into F . This completes the proof.

40. Prove Th. 5.2.1.19.

PROOF:

Since $x^2 + 1$ is irreducible over the real number field \bar{R} , the quotient field $\bar{R}/\{x^2 + 1\}$ is, by Prob. 38 above, to form a field.

Furthermore, since $x^2 + 1$ is a quadratic polynomial, every element of $\bar{R}[x]$ is congruent modulo $x^2 + 1$ to a linear polynomial, say $rx + s$, $r, s \in \bar{R}$, uniquely. Hence the elements of the quotient field $\bar{R}(x)/\{x^2 + 1\}$ is a residue class $\{rx + s\}$.

Since $\{rx + s\} = \{rx\} + \{s\} = \{r\} \cdot \{x\} + \{s\} \pmod{(x^2 + 1)}$, by Df. 4.1.2.3.18, let $\{r\} \leftrightarrow r$, $\{x\} \leftrightarrow x$, and $\{s\} \leftrightarrow s$, as in Prob. 39 above. Through this procedure it is now possible to express each element of the quotient field uniquely in the form of $rx + s$, of which two binary operations may be defined as follows:

$$\begin{aligned} (r_1x + s_1) + (r_2x + s_2) &= (r_1 + r_2)x + (s_1 + s_2) \\ (r_1x + s_1) \cdot (r_2x + s_2) &= r_1r_2x^2 + (r_1s_2 + r_2s_1)x + s_1s_2 \end{aligned} \quad (1)$$

Since $\{x^2 + 1\} = \{0\} \pmod{(x^2 + 1)}$ in this context, i.e. $x^2 + 1 = 0$ after the 1-1 mapping prescribed above, it follows that $x^2 = -1$, which in turn implies

$$(r_1x + s_1) \cdot (r_2x + s_2) = (s_1s_2 - r_1r_2) + (r_1s_2 + r_2s_1)x$$

Replace, then, x by i , and the operative rules of (1) with respect to the quotient field are readily mapped, through $r_1 \leftrightarrow b$, $r_2 \leftrightarrow a$, $s_1 \leftrightarrow d$, $s_2 \leftrightarrow c$, into:

$$\begin{aligned} (a + ib) + (c + id) &= (a + c) + i(b + d), \\ (a + ib) \cdot (c + id) &= (ac - bd) + i(ad + bc) \end{aligned}$$

or, what is the same,

$$\begin{aligned} (a, b) + (c, d) &= (a + c, b + d), \\ (a, b) \cdot (c, d) &= (ac - bd, ad + bc) \end{aligned} \quad (2)$$

which is strictly in accordance with the operative rules for the complex number field C (cf. Df. 5.1.3.1). Hence, by (1) and (2), the quotient field $\bar{R}(x)/\{x^2 + 1\}$ is now proved to be isomorphic to C , completing the proof.

41. If a polynomial $a(x)$ of positive degree is an element of a domain $F[x]$ of polynomials over a field F , then a field F' which contains F exists such that $a(x)$ has a root in F' .

PROOF:

If there exists no root in F of $a(x)$, then let $F' = F[x]/\{b(x)\}$, where $b(x)$ is a factor of $a(x)$, which is of degree at least two and is prime over F (cf. Th. 5.2.1.13). If, as in Prob. 40, the residue class $\{x\}$ is replaced by i , where $i \in F'$, then $b(i) = 0$, which in turn implies that the element i of F' is a root of $b(x)$, hence a root of $a(x)$, which may be considered a polynomial over F' in this context, since $F \subset F'$. A field containing F thus exists, in which $a(x)$ has a root.

If $a(x)$ has a root in F from the very beginning, the theorem is trivially true, since it merely implies $F' = F$.

Note. If $a(x)$ contains a factor of degree at least two and irreducible over F' , then the process prescribed above may be repeated, viz. constructing a field F'' , where $F' \subset F''$, and in general, a field $F^{(n)}$, where $F \subset F^{(n)}$ and $a(x)$ dissolves in $F^{(n)}(x)$ into linear factors.

42. Prove Th. 5.2.1.20.

PROOF:

Th. 5.2.1.7 has already established that there exist $b(x), c(x) \in F[x]$ such that

$$b(x)g(x) + c(x)f(x) = 1 \quad (1)$$

if $f(x), g(x) \in F[x]$ and $(f(x), g(x)) = 1$. Since evidently $f(x) \neq 0$ and $g(x) \neq 0$ in this context, divide (1) by $f(x)g(x)$ and multiply (1) by $a(x) \in F[x]$, yielding

$$a(x)/(f(x)g(x)) = (a(x)b(x))/f(x) + (a(x)c(x))/g(x)$$

which completes the proof.

43. If the denominator of a rational form $a(x)/b(x)$ is expressible as $(d(x))^r$, $r = 0, 1, 2, \dots, n$, then the rational form is expressible as a sum of partial fractions of the form $c_{n-r}(x)/(d(x))^r$, where $\deg c_r(x) < \deg(d(x))^r$.

PROOF:

Since $b(x) = (d(x))^r$ by hypothesis, divide $a(x)$ first by $d(x)$, which yields, by the division algorithm,

$$a(x) = q_0(x)d(x) + r_0(x) \quad (1)$$

If $\deg q_0(x) \geq \deg d(x)$ in (1), then divide $q_0(x)$ by $d(x)$ again, and

$$q_0(x) = q_1(x)d(x) + r_1(x) \quad (2)$$

Combine (1) and (2), and

$$a(x) = q_1(x)(d(x))^2 + r_1(x)d(x) + r_0(x) \quad (3)$$

If $\deg q_1(x) \geq \deg d(x)$ in (3), then the same process may be repeated again and again until, by induction, $a(x)$ is finally of the following form

$$a(x) = q_{n-1}(x)(d(x))^n + r_{n-1}(x)(d(x))^{n-1} + \dots + r_1(x)d(x) + r_0$$

which may be rewritten, replacing $q_{n-1}(x)$ by $c_n(x)$, $r_{n-r}(x)$ by $c_{n-r}(x)$ in general, and r_0 by $c_0(x)$, as follows:

$$a(x) = c_n(x)(d(x))^n + c_{n-1}(x)(d(x))^{n-1} + \dots + c_1(x)d(x) + c_0(x) \quad (4)$$

Divide (4) by $b(x) = (d(x))^n$, and let $(d(x))^0 = 1$; then

$$\begin{aligned} a(x)/b(x) &= c_n(x)/(d(x))^0 + c_{n-1}(x)/d(x) + \dots + c_1(x)/(d(x))^{n-1} + c_0(x)/(d(x))^n \\ &= \sum_r c_{n-r}(x)/(d(x))^r, \quad r = 0, 1, \dots, n \end{aligned} \quad (5)$$

which completes the proof.

44. Prove Th. 5.2.1.21.

PROOF:

To generalize Prob. 43, decompose $b(x)$ of the rational form $a(x)/b(x)$, as above, into a product of monic prime polynomials, and, combining equal prime polynomials together,

$$b(x) = b_0(p_1(x))^{n_1}(p_2(x))^{n_2} \dots (p_r(x))^{n_r} \quad (1)$$

where $n_k \in N$, $k=1,2,\dots,r$. Since any two distinct monic prime polynomials in (1), say $p_i(x)$ and $p_j(x)$, $i, j=1,2,\dots,r$, are relatively prime, so are their powers $(p_i(x))^{n_i}$ and $(p_j(x))^{n_j}$.

Now, let one of the prime factors of (1) above, say $(p_i(x))^{n_i}$, play the role of $(d(x))^n$ in the initial context of Prob. 43 to accomplish the factorization of $b(x)$. Then $a(x)/b(x)$ will be rendered in terms of partial fractions, each with a denominator of the form $(p_j(x))^{n_j}$, through the procedure similar to (3) and (4) in Prob. 43. The final step, similar to (5) in Prob. 43, will then complete the proof.

45. Decompose $f(x) = (x^5 + 4x^3 + x^2 + 5x - 3)/(x^4 - 1)$ into partial fractions.

Solution:

Taking the step (1) of Prob. 43,

$$f(x) = x + (4x^3 + x^2 + 6x - 3)/(x^4 - 1)$$

Since $x^4 - 1 = (x-1)(x+1)(x^2+1)$, it follows, by Prob. 43, that

$$\begin{aligned} (4x^3 + x^2 + 6x - 3)/(x^4 - 1) &= A/(x-1) + B/(x+1) + (Cx+D)/(x^2+1) \\ \text{i.e. } 4x^3 + x^2 + 6x - 3 &= A(x+1)(x^2+1) + B(x-1)(x^2+1) + (Cx+D)(x-1)(x+1) \end{aligned} \quad (1)$$

Substitute $x=1$ in (1); then $4A=8$, i.e. $A=2$. Substitute, likewise, $x=-1$ in (1); then $-4B=-12$, i.e. $B=3$. Substitute these results in (1); then $Cx+D=-x+2$. Hence

$$f(x) = x + 2/(x-1) + 3/(x+1) - (x-2)/(x^2+1)$$

Second Solution. From (1) above,

$$4x^3 + x^2 + 6x - 3 = (A+B+C)x^3 + (A-B+D)x^2 + (A+B-C)x + (A-B-D)$$

which implies

$$A+B+C=4, \quad A-B+D=1, \quad A+B-C=6, \quad A-B-D=-3$$

Solving these linear equations of order 4 simultaneously, we obtain $A=2$, $B=3$, $C=-1$, $D=2$. (Cf. Prob. 46, 47 below.)

Third Solution. Substitute $x=i$ in (1); then $-4i-1+6i-3=-2(Ci+D)$, which implies $i-2=-Ci-D$, i.e. $C=-1$ and $D=2$. This, together with the substitutions $x=\pm 1$ which yield $A=2$ and $B=3$, brings forth the same result. (Cf. Prob. 48 below.)

46. Decompose $f(x) = (x^3 - 2x^2 + 3x - 5)/((x-2)(x-1)^3)$ into partial fractions.

Solution:

By Th. 5.2.1.21,

$$\begin{aligned} f(x) &= A/(x-2) + B/(x-1)^3 + C/(x-1)^2 + D/(x-1) \\ \text{i.e. } x^3 - 2x^2 + 3x - 5 &= A(x-1)^3 + (x-2)(B+C(x-1)+D(x-1)^2) \end{aligned} \quad (1)$$

Substituting $x=2$ in (1), we obtain $A=1$. Substituting $A=1$ in (1) and simplifying,

$$x+2 = B+C(x-1)+D(x-1)^2 \quad (2)$$

Substitute $x=1$ in (2); then $B=3$. Furthermore, comparing the coefficients of x^2 and x , it follows that $D=0$ and $1=C-2D$, i.e. $C=1$. Hence

$$f(x) = 1/(x-2) + 3/(x-1)^3 + 1/(x-1)^2$$

47. Express $f(x) = 1/(x^4+1)$ as a sum of partial fractions over the real number field \bar{R} .

Solution:

Since x^4+1 is reducible to $(x^2+\sqrt{2}x+1)(x^2-\sqrt{2}x+1)$ over \bar{R} , it follows from Prob. 43 that

$$f(x) = (Ax+B)/(x^2+\sqrt{2}x+1) + (Cx+D)/(x^2-\sqrt{2}x+1) \quad (1)$$

i.e., after simplification,

$$(A+C)x^3 + (-\sqrt{2}A+B+\sqrt{2}D)x^2 + (A-\sqrt{2}B+C+\sqrt{2}D)x + (B+D) = 1 \quad (2)$$

which yields, after comparing coefficients,

$$A = 1/2\sqrt{2}, \quad B = 1/2, \quad C = -1/2\sqrt{2}, \quad D = 1/2$$

which thus allows $f(x)$ to decompose itself into

$$f(x) = (x + \sqrt{2})/2\sqrt{2} (x^2 + \sqrt{2}x + 1) - (x - \sqrt{2})/2\sqrt{2} (x^2 - \sqrt{2}x + 1)$$

Second Solution. Substitute $x = -x$ in (1) above, and

$$f(-x) = (-Ax + B)/(x^2 - \sqrt{2}x + 1) + (-Cx + D)/(x^2 + \sqrt{2}x + 1)$$

Compare (3) with (1), and it follows that, since the decomposition of $f(x)$ must be unique, $C = -A$ and $D = B$, which implies

$$f(x) = (Ax + B)/(x^2 + \sqrt{2}x + 1) - (Ax - B)/(x^2 - \sqrt{2}x + 1)$$

$$\text{i.e.} \quad (Ax + B)(x^2 - \sqrt{2}x + 1) - (Ax - B)(x^2 + \sqrt{2}x + 1) = 1 \quad (4)$$

Substitute $x = 0$ in (4), and $B = 1/2$ results; compare the coefficients of x^2 terms, and $-2\sqrt{2}A + 2B = 0$, i.e. $A = 1/2\sqrt{2}$, results, yielding the same decomposition as above.

48. Decompose $f(x) = 1/((x-1)^2(x^2+1)^2)$ over \bar{R} .

Solution:

By Th. 5.2.1.21,

$$f(x) = A/(x-1)^2 + B/(x-1) + (Cx + D)/(x^2 + 1)^2 + (Ex + F)/(x^2 + 1)$$

$$\text{i.e.} \quad (x^2 + 1)^2 (A + B(x-1)) + (x-1)^2 ((Cx + D) + (Ex + F)(x^2 + 1)) = 1 \quad (1)$$

Substitute $x = 1$ in (1), and $A = 1/4$ results; substitute $x = i$ in (1), and

$$(i-1)^2 (Ci + D) = 2C - 2Di = 1$$

from which $C = 1/2$ and $D = 0$. Substitute these values in (1), and

$$4B(x^2 + 1) + 4(Ex + F)(x-1) = -x - 3 \quad (2)$$

where $x = 1$ is substituted to yield $B = -1/2$ and $x = i$ is likewise substituted to yield $E = 1/2$ and $F = 1/4$. Hence

$$f(x) = 1/4(x-1)^2 - 1/2(x-1) + x/2(x^2+1)^2 + (2x+1)/4(x^2+1)$$

§5.2.2 Symmetric Polynomials

Df. 5.2.2.1 A polynomial $f(x_1, x_2, \dots, x_n)$ is called *symmetric* if it remains unchanged by any of the $n!$ permutations of the indeterminates x_1, x_2, \dots, x_n .

Example:

$abc(a+b+c)$, $a^2+b^2+c^2-ab-bc-ca$, $a^3+b^3+c^3+3abc$, $(a^4+b^4+c^4) - (a+b+c)^4$, etc., are symmetric polynomials with respect to a, b, c . (Cf. Prob. 1-2.)

It is evident in the context of Df. 3.1.2.8-9 that Df. 5.2.2.1 may have an alternative form:

Df. 5.2.2.1a A polynomial $f(x_1, x_2, \dots, x_n)$ is symmetric if it is invariant under the symmetric group of all permutations of its subscripts. (Cf. Prob. 17.)

Some symmetric polynomials are of specific forms, defined as follows:

Df. 5.2.2.2 The elementary symmetric polynomials are:

$$\begin{aligned} s_1(x_1, x_2, \dots, x_n) &= x_1 + x_2 + \dots + x_n, \\ s_2(x_1, x_2, \dots, x_n) &= x_1x_2 + x_1x_3 + \dots + x_1x_n + x_2x_3 + \dots + x_2x_n + \dots + x_{n-1}x_n, \\ s_3(x_1, x_2, \dots, x_n) &= x_1x_2x_3 + x_1x_2x_4 + \dots + x_{n-2}x_{n-1}x_n, \\ &\dots\dots\dots \\ s_i(x_1, x_2, \dots, x_n) &= x_1x_2 \dots x_i + \dots + x_{n-i+1}x_{n-1+2} \dots x_n, \\ &\dots\dots\dots \\ s_n(x_1, x_2, \dots, x_n) &= x_1x_2 \dots x_n. \end{aligned}$$

Example:

$a + b + c$, $ab + bc + ca$, and abc are the elementary symmetric polynomials of a, b, c .

Since $s_i(x_1, x_2, \dots, x_n)$ represents the sum of all the products formed by multiplying any i of the indeterminates x_1, x_2, \dots, x_n , it follows at once that s_i is symmetric with respect to x_1, x_2, \dots, x_n . For, since s_i is the sum of the products of the n indeterminates taken i at a time, it is the same as the sum of the products of $x_{j_1}, x_{j_2}, \dots, x_{j_n}$ taken i at a time, where j_1, j_2, \dots, j_n are the numbers $1, 2, \dots, n$ in some order.

As can be readily verified (cf. Th. 5.2.3.4), the elementary symmetric polynomials of Df. 5.2.2.2 reveals an important relation between the roots and coefficients of a polynomial, viz.,

$$(-1)^i s_i = S_i = (-1)^i a_{n-i}/a_n, \quad i = 1, 2, \dots, n$$

if x_1, x_2, \dots, x_n are the roots of the equation

$$f(y) = a_n y^n + a_{n-1} y^{n-1} + \dots + a_1 y + a_0 = 0$$

since, by hypothesis,

$$f(y) = a_n(y-x_1)(y-x_2) \dots (y-x_n) = a_n y^n + S_1 y^{n-1} + S_2 y^{n-2} + \dots + S_n$$

The meaning of the “elementary” symmetric polynomials is quite self-explanatory in the so-called Fundamental Theorem on symmetric polynomials:

Th. 5.2.2.3 Every symmetric polynomial in x_1, x_2, \dots, x_n over a field F can be written as a polynomial over F in the elementary symmetric polynomials of Df. 5.2.2.2: s_1, s_2, \dots, s_n . (Cf. Prob. 11 and Prob. 13.)

Example:

$a^3 + b^3 + c^3$, which is of course symmetric in a, b, c , can be expressed in terms of the elementary symmetric polynomials of Df. 5.2.2.2, viz.

$$a^3 + b^3 + c^3 = s_1^3 - 3s_1s_2 + 3s_3 = -S_1^3 + 3S_1S_2 - 3S_3$$

(Cf. Prob. 8 and Prob. 12.)

Th. 5.2.2.4 If f and g are symmetric polynomials in x_1, x_2, \dots, x_n , so are $f + g$, fg , and cf (or cg), where c is a constant. (Cf. Prob. 3.)

Example:

$x_1 + x_2 + x_3$, $x_1x_2 + x_1x_3 + x_2x_3$, and $x_1x_2x_3$ yield other symmetric polynomials such as $(x_1 + x_2 + x_3) + x_1x_2x_3$, $(x_1 + x_2 + x_3)x_1x_2x_3$, $c(x_1x_2 + x_1x_3 + x_2x_3)$, $cx_1x_2x_3(x_1 + x_2 + x_3)(x_1x_2 + x_1x_3 + x_2x_3)$, etc.

As is obvious in this example, addition and multiplication can be repeated, and Th. 5.2.2.4 may be stated more generally, readily justified by induction:

Th. 5.2.2.5 If f_1, f_2, \dots, f_m are symmetric polynomials in x_1, x_2, \dots, x_n , so are their sums and products. (Cf. Prob. 4.)

Generalization in a different direction yields:

Th. 5.2.2.6 If f_1, f_2, \dots, f_m are symmetric polynomials in x_1, x_2, \dots, x_n , and if g is a polynomial in y_1, y_2, \dots, y_m , then $g(f_1, f_2, \dots, f_m) = h(x_1, x_2, \dots, x_n)$ is a symmetric polynomial in x_1, x_2, \dots, x_n . (Cf. Prob. 6.)

Symmetric polynomials play a significant role with respect to the concept of functional independence, defined as follows:

Df. 5.2.2.7 The polynomials f_1, f_2, \dots, f_m in x_1, x_2, \dots, x_n are said to be *functionally dependent* over a field F if there exists a polynomial $g(y_1, y_2, \dots, y_m) \neq 0$ in F such that

$$g(f_1, f_2, \dots, f_m) = 0$$

Otherwise, the polynomials are said to be *functionally independent* over F .

Example:

$f_1 = x_1x_2$, $f_2 = x_1^2 + x_2^2$, and $f_3 = x_1^2 - x_2^2$ are functionally dependent, since $4f_1^2 - f_2^2 + f_3^2 = 0$, viz. there exists, in this context, $g(f_1, f_2, f_3) = 4y_1^2 - y_2^2 + y_3^2 = 0$.

Th. 5.2.2.8 The elementary symmetric polynomials of n indeterminates are functionally independent over any field which contains none of the indeterminates. (Cf. Prob. 16.)

Solved Problems

1. Express the symmetric polynomial

$$f(x, y, z) = (xy + z)(yz + x)(zx + y)$$

in terms of elementary symmetric polynomials.

Solution:

Multiply $f(x, y, z)$ by xyz , and by Df. 5.2.2.2,

$$\begin{aligned} xyz \cdot f(x, y, z) &= (xyz + z^2)(zyx + x^2)(zxy + y^2) = (s_3 + x^2)(s_3 + y^2)(s_3 + z^2) \\ &= s_3^3 + (x^2 + y^2 + z^2)s_3^2 + (x^2y^2 + y^2z^2 + z^2x^2)s_3 + x^2y^2z^2 \end{aligned}$$

where, however,

$$\begin{aligned} x^2 + y^2 + z^2 &= (x + y + z)^2 - 2(xy + yz + zx) = s_1^2 - 2s_2 \\ x^2y^2 + y^2z^2 + z^2x^2 &= (xy + yz + zx)^2 - 2(x + y + z)xyz = s_2^2 - 2s_1s_3 \end{aligned}$$

Hence

$$s_3 f(x, y, z) = s_3^3 + (s_1^2 - 2s_2)s_3^2 + (s_2^2 - 2s_1s_3)s_3 + s_3^2$$

which implies

$$f(x, y, z) = s_3^2 + (s_1^2 - 2s_2)s_3 + (s_2^2 - 2s_1s_3) + s_3$$

2. Factor $x^3 + y^3 + z^3 - (x + y + z)^3$.

Solution:

Let $f(x, y, z) = x^3 + y^3 + z^3 - (x + y + z)^3$; then $f(x, y, z) = 0$ if $x = -y$. Hence $(x + y) \mid f(x, y, z)$, and consequently, $(y + z) \mid f(x, y, z)$ and $(z + x) \mid f(x, y, z)$, since $f(x, y, z)$ is symmetric in x, y, z .

Furthermore, let

$$x^3 + y^3 + z^3 - (x + y + z)^3 = k(x + y)(y + z)(z + x) \quad (1)$$

Then, since both sides of the identity (1) are of the same degree, k must be a constant, which is quickly determined by substituting $x = y = 1$ and $z = 0$ in (1), viz. $k = -3$. Hence

$$f(x, y, z) = -3(x + y)(y + z)(z + x)$$

which is the desired result.

3. Prove Th. 5.2.2.4.

PROOF:

If i_1, i_2, \dots, i_n represent arbitrary rearrangements of $1, 2, \dots, n$, then by hypothesis,

$$\begin{aligned} f(x_{i_1}, x_{i_2}, \dots, x_{i_n}) &= f(x_1, x_2, \dots, x_n) \\ g(x_{i_1}, x_{i_2}, \dots, x_{i_n}) &= g(x_1, x_2, \dots, x_n) \end{aligned} \quad (1)$$

which together imply

$$f(x_{i_1}, x_{i_2}, \dots, x_{i_n}) + g(x_{i_1}, x_{i_2}, \dots, x_{i_n}) = f(x_1, x_2, \dots, x_n) + g(x_1, x_2, \dots, x_n) = h(x_1, x_2, \dots, x_n)$$

proving that $f + g$ is symmetric.

The product $f \cdot g$ is proved likewise, by the identities of (1), to be symmetric.

If, in particular, $g(x_1, x_2, \dots, x_n) = c$ (or $f(x_1, x_2, \dots, x_n) = c'$), it follows at once from the first part of the proof that cf (or $c'g$) is symmetric.

4. Prove Th. 5.2.2.5.

PROOF:

Since Th. 5.2.2.4 has already proved the case for $n = 2$, assume that Th. 5.2.2.5 is valid up to the case $n = k$, viz.,

$$\begin{aligned} f_1(x_{i_1}, x_{i_2}, \dots, x_{i_n}) + f_2(x_{i_1}, x_{i_2}, \dots, x_{i_n}) + \dots + f_k(x_{i_1}, x_{i_2}, \dots, x_{i_n}) \\ = f_1(x_1, x_2, \dots, x_n) + f_2(x_1, x_2, \dots, x_n) + \dots + f_k(x_1, x_2, \dots, x_n) \\ = g(x_1, x_2, \dots, x_n) \end{aligned}$$

which is symmetric in x_1, x_2, \dots, x_n . Then, by Th. 5.2.2.4,

$$g(x_{i_1}, x_{i_2}, \dots, x_{i_n}) + f_{k+1}(x_{i_1}, x_{i_2}, \dots, x_{i_n}) = g(x_1, x_2, \dots, x_n) + f_{k+1}(x_1, x_2, \dots, x_n)$$

which is again symmetric. Hence the proof is complete by induction.

The case of the products is proved likewise.

5. Factor $(x + y + z)^5 - (x^5 + y^5 + z^5)$.

Solution:

Let $f(x, y, z) = (x + y + z)^5 - (x^5 + y^5 + z^5)$, which is evidently a homogeneous symmetric polynomial of degree 5. As in Prob. 2 above, $f(x, y, z)$ is readily found to have a factor $(x + y)$ and, consequently, also $(y + z)$ and $(z + x)$. Hence $f(x, y, z)$ has a factor $(x + y)(y + z)(z + x)$, which is symmetric, of course. Hence, by Th. 5.2.2.4, it follows that the remaining factor of $f(x, y, z)$ must be also a homogeneous symmetric polynomial of degree 2, viz.,

$$\begin{aligned} f(x, y, z) &= (x + y + z)^5 - x^5 - y^5 - z^5 \\ &= (x + y)(y + z)(z + x)(a(x^2 + y^2 + z^2) + b(xy + yz + zx)) \end{aligned} \quad (1)$$

where a and b are constants. Substitute $x = y = 1, z = 0$, and also $x = y = z = 1$ in (1), and

$$30 = 2(2a + b), \quad 240 = 8(3a + 3b)$$

respectively. Solving them simultaneously for a and b , $a = b = 5$. Substituting in (1) and simplifying,

$$(x + y + z)^5 - (x^5 + y^5 + z^5) = 5(x + y)(y + z)(z + x)(x^2 + y^2 + z^2 + xy + yz + zx)$$

6. Prove Th. 5.2.2.6.

PROOF:

By hypothesis,

$$g(y_1, y_2, \dots, y_m) = \sum a_{i_1 i_2 \dots i_m} (y_1^{i_1} y_2^{i_2} \dots y_m^{i_m})$$

where the a 's are complex numbers and the summation notation denotes the sum of all the indicated products, the i 's being any non-negative integers subject to the condition: $0 \leq i_1 + i_2 + \dots + i_m \leq m$. Then, likewise,

$$h(x_1, x_2, \dots, x_n) = \sum a_{i_1 i_2 \dots i_m} (f_1^{i_1} f_2^{i_2} \dots f_m^{i_m})$$

which implies, by Th. 5.2.2.4-5, that each term of the summation, i.e. $a_{i_1 i_2 \dots i_m} (f_1^{i_1} f_2^{i_2} \dots f_m^{i_m})$, is symmetric. Hence, again by Th. 5.2.2.5, the sum of all the symmetric polynomial terms is also symmetric, completing the proof.

7. If f_0, f_1, \dots, f_m are polynomials in x_1, x_2, \dots, x_{n-1} , and if

$$g(x_1, x_2, \dots, x_{n-1}, x_n) = f_0 + x_n f_1 + \dots + x_n^m f_m$$

is a symmetric polynomial in x_1, x_2, \dots, x_n , then each of the f_i , $i = 0, 1, \dots, m$, is symmetric in x_1, x_2, \dots, x_{n-1} .

PROOF:

Since g is symmetric in x_1, x_2, \dots, x_n , by hypothesis, it follows, as in Prob. 4 above, that

$$g(x_1, x_2, \dots, x_{n-1}, x_n) \equiv g(x_{i_1}, x_{i_2}, \dots, x_{i_{n-1}}, x_n) \quad (1)$$

where i_1, i_2, \dots, i_{n-1} is an arbitrary arrangement of $1, 2, \dots, n-1$ just as $i_1, i_2, \dots, i_{n-1}, n$ is an arbitrary arrangement of $1, 2, \dots, n$. Hence, by (1) and hypothesis,

$$\begin{aligned} g(x_1, x_2, \dots, x_{n-1}, x_n) &= f_0 + x_n f_1 + \dots + x_n^m f_m \\ &= f_0(x_1, x_2, \dots, x_{n-1}) + x_n f_1(x_1, x_2, \dots, x_{n-1}) + \dots \\ &\quad + x_n^m f_m(x_1, x_2, \dots, x_{n-1}) \\ &= f_0(x_{i_1}, x_{i_2}, \dots, x_{i_{n-1}}) + x_n f_1(x_{i_1}, x_{i_2}, \dots, x_{i_{n-1}}) + \dots \\ &\quad + x_n^m f_m(x_{i_1}, x_{i_2}, \dots, x_{i_{n-1}}) \end{aligned}$$

which implies, by Th. 5.2.2.5, that each of

$$f_i(x_{i_1}, x_{i_2}, \dots, x_{i_{n-1}}) = f_i(x_1, x_2, \dots, x_{n-1}), \quad i = 0, 1, \dots, m$$

is symmetric, completing the proof.

Note. The second hypothesis is not limited to this problem; as a matter of fact, it is a theorem which directly follows from the definition of polynomials in n determinates (cf. Supplementary Problem 4.28-29).

*8. Express the sums of powers

$$P_m = x_1^m + x_2^m + \dots + x_n^m, \quad m = 1, 2, \dots, n-1$$

in the elementary symmetric polynomials of Df. 5.2.2.2.

Solution:

Since, by Df. 5.2.2.2,

$$f(x) = (x - x_1)(x - x_2) \cdots (x - x_n) = x^n + S_1 x^{n-1} + S_2 x^{n-2} + \dots + S_n \quad (1)$$

where $S_i = (-1)^i s_i$, $i = 1, 2, \dots, n$, s_i being the elementary symmetric polynomials, it follows from the Calculus that

$$f'(x)/f(x) = 1/(x - x_1) + 1/(x - x_2) + \dots + 1/(x - x_n) \quad (2)$$

where $f'(x)$ is the first derivative of $f(x)$. Rewrite (2), and

$$f'(x) = f(x)/(x - x_1) + f(x)/(x - x_2) + \dots + f(x)/(x - x_n) \quad (3)$$

and, by actual division, each term of the right-hand side of the equation (3) is found to be of the form

$$f(x)/(x - x_i) = x^{n-1} + f_1(x_i)x^{n-2} + f_2(x_i)x^{n-3} + \dots + f_{n-1}(x_i), \quad i = 1, 2, \dots, n \quad (4)$$

where $f_1(x_i) = x_i + S_1$, $f_2(x_i) = x_i^2 + S_1 x_i + S_2$, and in general,

$$f_j(x_i) = x_i^j + S_1 x_i^{j-1} + S_2 x_i^{j-2} + \dots + S_n, \quad j = 1, 2, \dots, n-1 \quad (5)$$

which is the coefficient of x^{n-j-1} .

Replace x_i in (4) successively by each of x_1, x_2, \dots, x_n and, substituting all of the results in (3),

$$\begin{aligned} f'(x) &= nx^{n-1} + (P_1 + nS_1)x^{n-2} + (P_2 + P_1 S_1 + nS_2)x^{n-3} + \dots \\ &\quad + (P_j + P_{j-1} S_1 + P_{j-2} S_2 + \dots + nS_j)x^{n-j-1} + \dots \end{aligned} \quad (6)$$

while, by (1) and the Calculus,

$$f'(x) = nx^{n-1} + (n-1)S_1 x^{n-2} + (n-2)S_2 x^{n-3} + \dots + 2S_{n-2} x + S_{n-1} \quad (7)$$

Hence, by equating the coefficients of the like powers of x in (6) and (7),

$$\begin{aligned} P_1 + S_1 &= 0, \\ P_2 + P_1 S_1 + 2S_2 &= 0, \\ P_3 + P_2 S_1 + P_1 S_2 + 3S_3 &= 0, \\ &\dots\dots\dots \\ P_{n-1} + P_{n-2} S_1 + P_{n-3} S_2 + \dots + P_1 S_{n-2} + (n-1)S_{n-1} &= 0 \end{aligned} \quad (8)$$

which yields, e.g.,

$$\begin{aligned} P_1 &= -S_1, \\ P_2 &= S_1^2 - 2S_2, \\ P_3 &= -S_1^3 + 3S_1 S_2 - 3S_3, \\ P_4 &= S_1^4 - 4S_1^2 S_2 + 4S_1 S_3 + 2S_2^2 - 4S_4, \\ P_5 &= -S_1^5 + 5S_1^3 S_2 + 5S_1 S_4 - 5S_1^2 S_3 - 5S_1 S_2^2 + 5S_2 S_3 - 5S_5 \end{aligned}$$

and so on, iterating the process, up to P_{n-1} , each of which is now expressed in terms of the elementary symmetric polynomials, as was desired.

Note. The equations of (8) above are the so-called Newton's formulas in the theory of equations.

9. Given $f_n = (x^n + y^n + z^n)(x^n y^n + y^n z^n + z^n x^n) - x^n y^n z^n$, prove that f_1 divides f_n if n is odd.

PROOF:

If n is odd, then f_n vanishes at the substitution of $x = -y$. Hence $(x+y)(y+z)(z+x)$ divides f_n if n is odd, since f_n is evidently a symmetric polynomial of degree $3n$ in x, y, z . But, in particular, f_1 is a symmetric polynomial of degree 3 in x, y, z , which therefore must be of the form

$$f_1 = c(x+y)(y+z)(z+x)$$

where c is a nonzero constant. Hence $f_1 \mid f_n$, completing the proof.

10. Prove, by Prob. 9, that $1/(a+b+c) = (1/a) + (1/b) + (1/c)$ implies

$$1/(a^n + b^n + c^n) = (1/a)^n + (1/b)^n + (1/c)^n$$

if n is odd.

PROOF:

Since, by Prob. 9,

$$f_1(a, b, c)/(abc(a+b+c)) = 1/a + 1/b + 1/c - 1/(a+b+c)$$

it follows that $f_1(a, b, c) = 0$ if $1/a + 1/b + 1/c = 1/(a+b+c)$. But, likewise,

$$f_n(a, b, c)/(a^n b^n c^n (a^n + b^n + c^n)) = (1/a)^n + (1/b)^n + (1/c)^n - 1/(a^n + b^n + c^n)$$

where $f_n(a, b, c) = 0$ if n is odd, as has already been proved in Prob. 9. Hence, if n is odd,

$$1/(a^n + b^n + c^n) = (1/a)^n + (1/b)^n + (1/c)^n$$

11. Prove Th. 5.2.2.3.

PROOF:

(i) Limit the case, first, to two indeterminates x_1, x_2 , and their elementary symmetric functions are

$$s_1 = x_1 + x_2, \quad s_2 = x_1 x_2 \quad (1)$$

Given a symmetric polynomial in x_1, x_2 , say $f(x_1, x_2)$, it can be rearranged in powers of x_1 , yielding

$$f(x_1, x_2) = A_a x_1^a + A_{a-1} x_1^{a-1} + \dots + A_0 \quad (2)$$

where A_a, A_{a-1}, \dots, A_0 are polynomials in x_2 . If x_2 in (2) is replaced by $s_1 - x_1$, which is justified by (1), then $f(x_1, x_2)$ becomes a polynomial in x_1 alone, viz.,

$$f(x_1, x_2) = B_b x_1^b + B_{b-1} x_1^{b-1} + \cdots + B_0 = g(x_1) \quad (3)$$

where B_b, B_{b-1}, \dots, B_0 are polynomials in s_1 , which are all arranged in powers of x_1 .

Now divide $g(u) = B_b u^b + B_{b-1} u^{b-1} + \cdots + B_0$ by $h(u) = u^2 - s_1 u + s_2$, and then, by Th. 5.2.1.2,

$$g(u) = h(u) q(u) + Cu + D \quad (4)$$

where C and D are polynomials in s_1, s_2 in a manner that satisfies Th. 5.2.2.4-5. Since $g(x_1) = f(x_1, x_2)$ and $h(x_1) = 0$, $u = x_1$ in (4) implies

$$f(x_1, x_2) = Cx_1 + D$$

where x_1 and x_2 may be interchanged, since C and D are by definition not to be affected by such an interchange. Thus

$$Cx_1 + D = Cx_2 + D$$

which implies

$$C(x_1 - x_2) = 0$$

which in turn implies, since neither x_1 nor x_2 is zero, that C must be zero. Hence

$$f(x_1, x_2) = D$$

viz. the symmetric polynomial f in x_1, x_2 is now given in terms of their elementary symmetric polynomials.

- (ii) Since the case for two indeterminates is now verified, assume that the theorem is valid up to the case of $n-1$ indeterminates.

If t_1, t_2, \dots, t_{n-1} are the elementary symmetric polynomials of the $n-1$ indeterminates x_2, x_3, \dots, x_n , then

$$\begin{aligned} s_1 &= x_1 + t_1, \\ s_2 &= x_1 t_1 + t_2, \\ &\dots\dots\dots \\ s_{n-1} &= x_1 t_{n-2} + t_{n-1}, \end{aligned} \quad (5)$$

and conversely,

$$\begin{aligned} t_1 &= -x_1 + s_1, \\ t_2 &= x_1^2 - x_1 s_1 + s_2, \\ &\dots\dots\dots \\ t_{n-1} &= (-1)^{n-1} (x_1^{n-1} - x_1^{n-2} s_1 + \cdots + (-1)^{n-1} s_{n-1}) \end{aligned} \quad (6)$$

Let $f(x_1, x_2, \dots, x_n)$ be symmetric in x_1, x_2, \dots, x_n , and arrange it in powers of x_1 , as in (i), yielding

$$f(x_1, x_2, \dots, x_n) = A_a x_1^a + A_{a-1} x_1^{a-1} + \cdots + A_0 \quad (7)$$

where A_a, A_{a-1}, \dots, A_0 are symmetric polynomials in x_2, x_3, \dots, x_n , which can be expressed as polynomials in t_1, t_2, \dots, t_{n-1} , since the theorem has been assumed to be valid in the case of $n-1$ indeterminates.

Also, because of the relation between (5) and (6),

$$f(x_1, x_2, \dots, x_n) = B_b x_1^b + B_{b-1} x_1^{b-1} + \cdots + B_0 = g(x_1) \quad (8)$$

where B_b, B_{b-1}, \dots, B_0 are polynomials in s_1, s_2, \dots, s_{n-1} , since A_a, A_{a-1}, \dots, A_0 in (7) can be expressed as polynomials in $x_1, s_1, s_2, \dots, s_{n-1}$ while t_1, t_2, \dots, t_{n-1} are replaced by their expressions through $x_1, s_1, s_2, \dots, s_{n-1}$.

Divide, then, $g(u) = B_b u^b + B_{b-1} u^{b-1} + \cdots + B_0$ by $h(u) = u^n - s_1 u^{n-1} + s_2 u^{n-2} - \cdots + (-1)^n s_n$, and, by Th. 5.2.1.2,

$$g(u) = h(u) q(u) + C_{n-1} u^{n-1} + C_{n-2} u^{n-2} + \cdots + C_0 \quad (9)$$

where $C_{n-1}, C_{n-2}, \dots, C_0$ are polynomials in s_1, s_2, \dots, s_n which comply with Th. 5.2.2.4-5. Let $u = x_1$ in (9); then, since $h(x_1) = 0$, it follows that

$$f(x_1, x_2, \dots, x_n) = C_{n-1} x_1^{n-1} + C_{n-2} x_1^{n-2} + \cdots + C_0 \quad (10)$$

where x_1 and x_2, x_3, \dots, x_n may be interchanged as in (i), which implies the following identities in x_1, x_2, \dots, x_n :

Since both (2) and (3) express the homogeneous symmetric polynomial as a rational integral polynomial of the P_k , where $k \in N$, and since the Newton's formulas (cf. Prob. 8) yield the P_k as rational integral polynomials of the S 's, the theorem is now verified up to the case of symmetric polynomials in two indeterminates x_1, x_2 .

- (ii) Assume that the theorem holds up to symmetric polynomials in m indeterminates x_1, x_2, \dots, x_m ; viz. any homogeneous symmetric polynomial, each term of which involves m roots, can be represented as a rational integral polynomial of the P_k . Since, by this assumption,

$$\sum x_1^a x_2^b \cdots x_m^p = x_1^a x_2^b \cdots x_m^p + x_1^b x_2^a \cdots x_m^p + \cdots + x_1^p x_2^b \cdots x_m^a + \cdots$$

and

$$P_q = x_1^q + x_2^q + \cdots + x_m^q$$

where a, b, \dots, p are all distinct and also $q \neq a, b, \dots, p$, it follows that

$$P_q \sum x_1^a x_2^b \cdots x_m^p = \sum x_1^{a+q} x_2^b \cdots x_m^p + \sum x_1^a x_2^{b+q} \cdots x_m^p + \cdots + \sum x_1^a x_2^b \cdots x_m^{p+q} + \sum x_1^a x_2^b \cdots x_m^p x_{m+1}^q \quad (4)$$

revealing that the symmetric polynomial $\sum x_1^a x_2^b \cdots x_m^p x_{m+1}^q$ of $m+1$ roots can be expressed in terms of homogeneous symmetric polynomials, each term of which involves m roots, and one such symmetric polynomial multiplied by P_q . Hence the symmetric polynomial in $m+1$ indeterminates can be expressed as a rational integral polynomial of the P_k , and thus, by Prob. 8, as a rational integral polynomial of the coefficients of (1), completing the proof.

Note. The term "homogeneous" in the hypothesis of $\sum x_1^a x_2^b$ is not exactly essential, since any non-homogeneous rational integral symmetric polynomial is the sum of two or more homogeneous rational integral symmetric polynomials.

Furthermore, the theorem is readily made more general, as in Prob. 14 below.

14. If $f(x)$ is a polynomial of degree n over a field F with roots x_1, x_2, \dots, x_n , and if $g(y_1, y_2, \dots, y_n)$ is a symmetric polynomial over F , then $g(x_1, x_2, \dots, x_n)$ is an element of F .

PROOF:

Since, by Prob. 13, $g(y_1, y_2, \dots, y_n)$ is a polynomial over F in the elementary symmetric polynomials s_1, s_2, \dots, s_n , it follows that $g(x_1, x_2, \dots, x_n)$ is a polynomial in

$$x_1 + x_2 + \cdots + x_n, \quad x_1 x_2 + x_1 x_3 + \cdots + x_{n-1} x_n, \quad x_1 x_2 \cdots x_n$$

These expressions, however, are merely the coefficients of $f(x)/a_n$ if

$$f(x) = a_n(x^n - b_{n-1}x^{n-1} + b_{n-2}x^{n-2} - \cdots + (-1)^n b_0) \quad (1)$$

where evidently $b_k \in F$, $k = 0, 1, \dots, n-1$. Since the expression (1) always holds, the proof is complete.

15. If $a(x)$ and $b(x)$ are polynomials over a field F with a_1, a_2, \dots, a_m and b_1, b_2, \dots, b_n as their respective roots, then the products

$$p(x) = \prod_i \prod_j (x - (a_i + b_j)), \quad i = 1, 2, \dots, m$$

$$q(x) = \prod_i \prod_j (x - (a_i b_j)), \quad j = 1, 2, \dots, n$$

are polynomials in x with coefficients in F .

PROOF:

- (i) Since, by hypothesis,

$$a(x) = c_m(x - a_1)(x - a_2) \cdots (x - a_m)$$

where c_m is the leading coefficient of $a(x)$, it follows that

$$a(x - b_j) = c_m(x - (a_1 + b_j))(x - (a_2 + b_j)) \cdots (x - (a_m + b_j)) = c_m \prod_i (x - (a_i + b_j))$$

Hence, by hypothesis,

$$c_m^n p(x) = \prod_j a(x - b_j) \quad (1)$$

which is a polynomial in x with coefficients which are symmetric in b_1, b_2, \dots, b_n . Thus, by Prob. 14, the coefficients of (1) must be in F . Further, if both sides of (1) are divided by c_m^n , it follows at once that the coefficients of $p(x)$ are in F as it is a field.

(ii) As is justifiable by hypothesis, let

$$a(x/b_j) = c_m((x/b_j) - a_1)((x/b_j) - a_2) \cdots ((x/b_j) - a_m)$$

which implies

$$b_j^m a(x/b_j) = c_m(x - a_1 b_j)(x - a_2 b_j) \cdots (x - a_m b_j)$$

which in turn implies

$$c_m^n q(x) = \prod_j b_j^m a(x/b_j) \quad (2)$$

Hence, as in (i), divide both sides of (2) by c_m^n , and the coefficients of $q(x)$ are found to be in F , completing the proof.

16. Prove Th. 5.2.2.8.

PROOF:

The proof is to show that

$$f(s_1, s_2, \dots, s_n) \neq 0 \quad (1)$$

where s_1, s_2, \dots, s_n are the elementary symmetric polynomials of the n indeterminates x_1, x_2, \dots, x_n ,

and

$$f(y_1, y_2, \dots, y_n) = \sum_{k_1, k_2, \dots, k_n} a_{k_1 k_2 \dots k_n} y_1^{k_1} y_2^{k_2} \cdots y_n^{k_n} \neq 0 \quad (2)$$

is a polynomial in a field F which contains none of x_1, x_2, \dots, x_n ; the polynomial is of the reduced form so that no two distinct terms of (2) are of the same powers of y_1, y_2, \dots, y_n .

Now let a term of (2),

$$a_{k_1 k_2 \dots k_n} y_1^{k_1} y_2^{k_2} \cdots y_n^{k_n} \quad (3)$$

be chosen by considering only those terms of (2) for which $k_1 + k_2 + \cdots + k_n$ is a maximum, then again those terms for which $k_2 + k_3 + \cdots + k_n$ is a maximum, and so on. These terms are unique, for the identities

$$\begin{aligned} k_1 + k_2 + \cdots + k_n &= k'_1 + k'_2 + \cdots + k'_n, \\ k_2 + \cdots + k_n &= k'_2 + \cdots + k'_n, \\ &\dots\dots\dots \\ k_n &= k'_n, \end{aligned}$$

will imply $k_1 = k'_1, k_2 = k'_2, \dots, k_n = k'_n$.

Since, by hypothesis and Df. 5.2.2.2,

$$s_1^{k_1} s_2^{k_2} \cdots s_n^{k_n} = (x_1 + x_2 + \cdots + x_n)^{k_1} (x_1 x_2 + x_1 x_3 + \cdots + x_{n-1} x_n)^{k_2} \cdots (x_1 x_2 \cdots x_n)^{k_n}$$

one of the terms which occurs when $a_{k_1 k_2 \dots k_n} s_1^{k_1} s_2^{k_2} \cdots s_n^{k_n}$ is expressed in terms of x_1, x_2, \dots, x_n is

$$a_{k_1 k_2 \dots k_n} x_1^{k_1} (x_1 x_2)^{k_2} \cdots (x_1 x_2 \cdots x_n)^{k_n} = a_{k_1 k_2 \dots k_n} x_1^{k_1 + k_2 + \cdots + k_n} x_2^{k_2 + \cdots + k_n} \cdots x_{n-1}^{k_{n-1} + k_n} x_n^{k_n} \quad (4)$$

which is unique in that no other term is of the same powers.

Under the maximal conditions imposed on the k 's, no term with the same exponents as those of the right member of (4) can appear when any other term of $f(s_1, s_2, \dots, s_n)$ is expressed in x_1, x_2, \dots, x_n . Thus the right member of (4) does come into existence when $f(s_1, s_2, \dots, s_n)$, expressed in x_1, x_2, \dots, x_n , is simplified by combining like terms. That is, $f(s_1, s_2, \dots, s_n) \neq 0$, completing the proof.

17. Given a polynomial $p = p(x_1, x_2, \dots, x_n)$, the set A of all permutations a , of $1, 2, \dots, n$, such that $p^a = p$ forms a group. So does the set B of all permutations b such that $p^b = \pm p$.

PROOF:

(i) Let $a_1, a_2, a_3 \in A$; then, by Df. 3.1.1.1,

G1. $a_1 a_2 \in A$, since $p^{a_1 a_2} = (p^{a_1})^{a_2} = p^{a_2} = p$.

G2. $a_1(a_2 a_3) = (a_1 a_2) a_3$, since $p^{a_1(a_2 a_3)} = p^{(a_1 a_2) a_3} = p$.

G3. $1 \in A$.

G4. $a^{-1} \in A$, since $p^a = p$ implies $(p^a)^{a^{-1}} = p^{a^{-1}}$, which in turn implies $p = p^{a^{-1}}$.

Hence A is a group.

(ii) Likewise, let $b_1, b_2, b_n \in B$; then

G1. $b_1 b_2 \in B$, since $p^{b_1 b_2} = (p^{b_1})^{b_2} = (\pm p)^{b_2} = \pm(\pm p) = \pm p$.

G2. $b_1(b_2 b_3) = (b_1 b_2)b_3$, since $p^{b_1(b_2 b_3)} = p^{(b_1 b_2)b_3} = \pm p$.

G3. $1 \in B$.

G4. $b^{-1} \in B$, since $p^b = \pm p$ implies $(p^b)^{b^{-1}} = \pm p^{b^{-1}}$, yielding $p^{b^{-1}} = \pm p$.

Hence B is a group.

§5.2.3 Roots of Polynomials

Th. 5.2.3.1 If a polynomial $f(x) = \sum_k a_k x^k$, $k=0,1,\dots,n$, vanishes (i.e. $f(x)=0$) for more than n distinct values of x , then it is identically zero (i.e. $a_0=a_1=\dots=a_n=0$). (Cf. Prob. 1.)

This theorem is a direct consequence of Th. 5.2.1.5 (and Th. 4.1.2.5.18) that $f(x)$ of degree n cannot have more than n roots. A polynomial, however, may not have any root at all; for example, a constant polynomial $f(x) \equiv a_0 x^0$ has no root if $a_0 \neq 0$, since it will never vanish for any value of x . On the other hand, $f(x)$ may have a root which occurs more than once, as is made explicit in the following definition (which has been implicitly used, e.g. in Th. 5.2.1.9, Th. 5.2.1.21, etc.).

Df. 5.2.3.2 If a polynomial $f(x)$ has a factor of the form $(x-r)^m$, but not $(x-r)^{m+1}$, then r is called a root of *multiplicity* m (or an *m-fold* root).

Example:

$f(x) = (x-r)^p g(x)$ has a p -fold root r which as such cannot be a root of $g(x)=0$; if $f(x)$ is a polynomial of degree n , then it cannot have more than $n-p+1$ distinct roots in this context, as is proved in the following theorem:

Th. 5.2.3.3 A polynomial of degree n cannot have more than n roots unless a root of multiplicity m , if any, is counted m times. (Cf. Prob. 3.)

The roots of a polynomial are related to the coefficients of the polynomial in a definite pattern, generalized as follows:

Th. 5.2.3.4 If r_1, r_2, \dots, r_n are the roots of the equation

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0 = 0 \quad (1)$$

then the sum of all possible products of the r 's taken k at a time, $k=1,2,\dots,n$, is equal to $(-1)^{n-k} a_{n-k}/a_n$. (Cf. Prob. 4.)

Stated in detail (cf. Df. 5.2.2.2): $f(x) = a_n(x-r_1)(x-r_2)\dots(x-r_n)$ yields

$$\begin{aligned} S_1 &= r_1 + r_2 + \dots + r_n = -a_{n-1}/a_n, \\ S_2 &= r_1 r_2 + r_1 r_3 + \dots + r_{n-1} r_n = a_{n-2}/a_n, \\ S_3 &= r_1 r_2 r_3 + r_1 r_2 r_4 + \dots + r_{n-2} r_{n-1} r_n = -a_{n-3}/a_n, \\ &\dots\dots\dots \\ S_n &= r_1 r_2 \dots r_n = (-1)^n a_0/a_n. \end{aligned} \quad (2)$$

Example:

$a_2x^2 + a_1x + a_0 = 0$ has two roots r_1 and r_2 such that $S_1 = r_1 + r_2 = -a_1/a_2$ and $S_2 = r_1r_2 = a_0/a_2$.

As has already been observed (cf. Df. 5.2.2.2), S_1, S_2, \dots, S_n in this context are readily expressible in terms of the elementary symmetric polynomials, s_1, s_2, \dots, s_n ; viz.

$$S_i = (-1)^i s_i, \quad i = 1, 2, \dots, n$$

The relation between the roots and coefficients of a polynomial helps solve an equation when a certain relation among the roots is given; it also helps construct an equation whose roots have an assigned relation to the roots of a given equation (cf. Prob. 5-10).

In general, for nonconstant polynomials at least, the following theorem holds:

Th. 5.2.3.5 If $f(a) < 0$ and $f(b) > 0$, where $f(x)$ is a polynomial over \bar{R} and $a, b \in \bar{R}$, then there exists $c \in \bar{R}$ between a and b such that $f(c) = 0$. (Cf. Prob. 11.)

This theorem, as a special case of the so-called Mean-value Theorem, holds for any continuous functions as well. It is, however, essentially an existence theorem, which as such does not pin-point the root of the given polynomial in an actual construction, although it may be employed as a principle for locating approximate values of the real roots of $f(x)$. It is of practical importance to be able to approximate the real roots with any required degree of accuracy, in particular for polynomials of fifth and higher degree, or even third and fourth degree; but the principles of approximation to real roots are definitely outside the legitimate scope of the present work.

Th. 5.2.3.5 yields also the following theorem:

Th. 5.2.3.6 Every polynomial with real coefficients of odd degree has at least one real root. (Cf. Prob. 12.)

Just as polynomials with real coefficients have their own peculiarities as above (and also as in Th. 5.2.1.13-14 and Prob. 27-36 below), a restriction of coefficients to the rational number field R or, what is essentially the same, to the domain I of integers yields the following simple theorems:

Th. 5.2.3.7 An integral root of a polynomial with integral coefficients is an exact divisor of the constant term of the polynomial. (Cf. Prob. 18.)

Th. 5.2.3.8 If a polynomial with integral coefficients has a rational root of the form p/q , where p and q are integers which have no common divisor other than unity, then p is an exact divisor of the constant term, and q an exact divisor of the leading coefficient, of the polynomial (Cf. Prob. 19.)

The restriction of coefficients of polynomials as above may be replaced by a generalization, extending the coefficients from R or \bar{R} to C (cf. Th. 5.2.1.15-16), and the following theorem, called *the Fundamental Theorem of Algebra*, is of course the most important for polynomials with complex coefficients.

Th. 5.2.3.9 (by Euler-Gauss). Every polynomial of positive degree with complex coefficients has a complex root.

This is again an existence theorem, which does not help find *in concreto* the roots of a polynomial, but does assert unequivocally that they do exist. The theorem

asserts, in fact, that every algebraic equation of degree greater than zero (i.e. a nonconstant polynomial) has at least one root, real or complex; the validity of this theorem has been taken for granted, without proof, throughout College Algebra. In the present context, too, the theorem will be presumed to be true.

(Contrary to the belief of some authors, however, it is not at all the case that the proof without any knowledge of function theory is "either tedious or lacking in rigor", although the theory of functions of a complex variable does offer a simpler proof. Nor should the student be misled to consider Th.5.2.3.9 the fundamental theorem of algebra in modern algebra, although the rôle the complex number field, or the theory of equations in general, plays is still unique and great in the same context.)

The following theorems are direct results from Th.5.2.3.9.

Th. 5.2.3.10 Polynomials with complex coefficients of degree n have n complex roots. (Cf. Th.5.2.3.3 and also Prob. 20 below.)

This theorem yields the following definition:

Df. 5.2.3.10a The complex number field in relation to Th.5.2.3.10 is said to be *algebraically closed* (or *complete*).

In this sense any field F may be said to be algebraically complete (or closed) if polynomials over F have no roots outside F itself.

Th. 5.2.3.11 If $f(z)=0$, where $z \in C$ and $f(x)$ is a polynomial with real coefficients, then $f(\bar{z})=0$. (Cf. Df.5.1.3.8 and Prob. 21 below.)

Th. 5.2.3.12 Every polynomial $f(x)$ with real coefficients of degree greater than one is expressible over \bar{R} as

$$f(x) = a_n(x-b_1) \cdots (x-b_r)((x-c_1)^2+d_1^2) \cdots ((x-c_s)^2+d_s^2)$$

where $a_n \neq 0$ is the leading coefficient of $f(x)$ and $b_i, c_j, d_j \in \bar{R}$, $i=1,2,\dots,r$, $j=1,2,\dots,s$, and $d_j \neq 0$. (Cf. Prob. 24.)

Th. 5.2.3.13 If $f(x)$ is a polynomial with real coefficients, and if $f(a) \neq 0$ and $f(b) \neq 0$, where $a, b \in \bar{R}$, then either $f(a)f(b) > 0$ or $f(a)f(b) < 0$ according as the number of the real roots of $f(x)$ between a and b is either even or odd. (Cf. Prob. 25.)

The last theorem belongs to the so-called *isolation* of the real roots of a polynomial with real coefficients, which examines whether one or more intervals can be found such that each real root is contained in one of these intervals and each interval contains only one root. This leads to such theorems as Sturm's and Budan's and also to Descartes' rule of signs.

Now that the existence of roots for nonconstant polynomials is assured, the problem is directed to the *algebraic* solution of individual equations, viz. to find the roots of polynomials by rational operations and radicals alone, through the following two definitions:

Df. 5.2.3.14 The equation $x^n=a$ is called a *binomial* equation, where the roots of the equation are the n th roots of a , denoted by $\sqrt[n]{a}$ and called a *radical of index n* (relative to a field which contains a). (Cf. Prob. 28-29 below.)

The radical, defined as above, leads to the following definition of solvability.

Df. 5.2.3.15 An equation $f(x) = \sum a_i x^i = 0$, $i=0,1,\dots,n$, is said to be *solvable by radicals* (or *root extractions*) and rational operations if there exists a sequence of numbers b_j , $j=1,2,\dots,m$, such that every b_j is either one of a_i or one of the results of rational operations on b_{j_r} and b_{j_s} , where $0 \leq b_{j_r}, b_{j_s} < m$, or a root of any index of a preceding b or one of the results of rational operations on b_{j_r} and b_{j_s} , where $1 \leq b_{j_r}, b_{j_s} < m$, such that every root of the equation appears in the sequence.

Example:

The sequence for $a_2 x^2 + a_1 x + a_0 = 0$, $a_2 \neq 0$, is:

$b_1 = a_2$	$b_6 = b_1 b_3 = a_2 a_0$	$b_{11} = b_{10} - b_9 = a_1^2 - 4a_2 a_0$
$b_2 = a_1$	$b_7 = b_4 + b_4 = 2$	$b_{12} = \sqrt{b_{11}} = \sqrt{a_1^2 - 4a_2 a_0}$
$b_3 = a_0$	$b_8 = b_7 + b_7 = 4$	$b_{13} = b_5 - b_{12} = -\sqrt{a_1^2 - 4a_2 a_0}$
$b_4 = b_1/b_1 = 1$	$b_9 = b_8 b_8 = 4b_8 = 4a_2 a_0$	etc.
$b_5 = b_1 - b_1 = 0$	$b_{10} = b_2 b_2 = a_1^2$	

There exist similar but different sequences for the given equation, as in most cases; at least one such sequence must exist, however, for any equation to be solvable by radicals.

In general, there exists at least one such sequence for every equation of positive degree up to 4, beyond which such sequences may or may not exist. Every equation of degree up to 4, then, will have a general formula involving only rational operations and root extractions, as will be seen below, for expressing its roots in terms of its coefficients; such a general formula does not exist, however, for any equation of degree greater than 4.

Th. 5.2.3.16 The linear equation, generally of the form $a_1 x + a_0 = 0$, where $a_1, a_0 \in C$ and $a_1 \neq 0$, has the unique root $-a_0/a_1$.

Th. 5.2.3.17 The quadratic equation, generally of the form $a_2 x^2 + a_1 x + a_0 = 0$, where $a_2, a_1, a_0 \in C$ and $a_2 \neq 0$, has two roots:

$$r_1 = (-a_1 + \sqrt{a_1^2 - 4a_2 a_0})/2a_2, \quad r_2 = (-a_1 - \sqrt{a_1^2 - 4a_2 a_0})/2a_2$$

The student is already familiar with these two results (cf. also §5.1.3, Prob. 28-29).

Th. 5.2.3.18 (by Cardano). The cubic equation, generally of the form

$$a_3 x^3 + a_2 x^2 + a_1 x + a_0 = 0$$

where $a_3, a_2, a_1, a_0 \in C$ and $a_3 \neq 0$, can be simplified to the form

$$\hat{x}^3 + p\hat{x} + q = 0$$

which can be solved algebraically. (Cf. Prob. 34.)

Th. 5.2.3.19 (by Ferrari-Euler). The quartic equation, generally of the form

$$a_4 x^4 + a_3 x^3 + a_2 x^2 + a_1 x + a_0 = 0$$

where $a_4, a_3, a_2, a_1, a_0 \in C$ and $a_4 \neq 0$, can be simplified to the form

$$\tilde{x}^4 + p\tilde{x}^2 + q\tilde{x} + r = 0$$

which can be algebraically solved. (Cf. Prob. 36.)

This theorem, however, is the end of algebraic solutions of general equations (i.e. without any made-to-fit modifications of coefficients). For, as is well-known,

Abel proved it once and for all that it is absolutely impossible to express the roots of an equation of degree higher than the fourth by means of formulas involving only rational operations and radicals, viz.:

Th. 5.2.3.20 (by Abel). The quintic equation, or any equation of degree higher than fifth, in its general form is not algebraically solvable.

It must be noted that the same result can be obtained by a theory originally conceived by Galois.

Solved Problems

1. Prove Th. 5.2.3.1.

PROOF:

Given $f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$, and assume $a_n \neq 0$. Then $f(x)$ has more than n distinct roots, by hypothesis, which is a flat contradiction in relation to Th. 5.2.1.5. Hence it must be the case that $a_n = 0$.

Assume, then, $a_{n-1} \neq 0$, which implies, by hypothesis, that $f(x)$ has more than $n-1$ distinct roots, which is again contradictory to Th. 5.2.1.5. Hence $a_{n-1} = 0$ must be the case.

Likewise, taking similar steps, $a_{n-2} = 0, \dots, a_1 = 0, a_0 = 0$, i.e. $a_n = a_{n-1} = \cdots = a_1 = a_0 = 0$, which completes the proof.

2. If $f(x) = \sum_k a_k x^k$ and $g(x) = \sum_k b_k x^k$, $k=0,1,\dots,n$, have the same value for more than n distinct values of x , then $f(x) = g(x)$, where $a_k = b_k$.

PROOF:

Let

$$h(x) = f(x) - g(x) = \sum_k a_k x^k - \sum_k b_k x^k = \sum_k (a_k - b_k) x^k = \sum_k c_k x^k, \quad k=0,1,\dots,n$$

Then, by Th. 5.2.3.1, $c_n = c_{n-1} = \cdots = c_0 = 0$, where $c_k = a_k - b_k$, which implies $a_n = b_n, a_{n-1} = b_{n-1}, \dots, a_0 = b_0$, completing the proof.

3. Prove Th. 5.2.3.3.

PROOF:

Let $f(x)$ be a polynomial of degree n with roots r_1, r_2, \dots, r_k of multiplicity m_1, m_2, \dots, m_k . Then, by Df. 5.2.3.2,

$$f(x) = a_n(x-r_1)^{m_1}(x-r_2)^{m_2} \cdots (x-r_k)^{m_k} g(x) \quad (1)$$

where a_n is the leading coefficient of $f(x)$ and evidently $\deg g(x) = n - (m_1 + m_2 + \cdots + m_k)$ by Th. 5.2.1.5. Since the expression (1) is unique, by Df. 5.2.3.2 itself, the proof is complete.

4. Prove Th. 5.2.3.4.

PROOF:

Since the cases $n=1,2$ are trivial, take the case $n=3$, and, by actual computation,

$$\begin{aligned} f(x) &= a_3 x^3 + a_2 x^2 + a_1 x + a_0 = a_3(x-r_1)(x-r_2)(x-r_3) \\ &= a_3(x^3 - (r_1+r_2+r_3)x^2 + (r_1r_2+r_1r_3+r_2r_3)x - r_1r_2r_3) = a_3(x^3 - A_1x^2 + A_2x - A_3) \end{aligned}$$

which implies

$$A_1 = r_1 + r_2 + r_3 = -a_2/a_3, \quad A_2 = r_1r_2 + r_1r_3 + r_2r_3 = a_1/a_3, \quad A_3 = r_1r_2r_3 = -a_0/a_3$$

Assume, then, that the case $n = k$ holds, viz.,

$$\begin{aligned} f(x) &= a_k x^k + a_{k-1} x^{k-1} + \cdots + a_0 = a_k(x-r_1)(x-r_2)\cdots(x-r_k) \\ &= a_k(x^k - (r_1 + r_2 + \cdots + r_k)x^{k-1} + (r_1r_2 + r_1r_3 + \cdots + r_{k-1}r_k)x^{k-2} - \cdots + (-1)^k r_1r_2\cdots r_k) \\ &= a_k(x^k - K_1x^{k-1} + K_2x^{k-2} - \cdots + (-1)^k K_k) \end{aligned}$$

which implies

$$\begin{aligned} K_1 &= r_1 + r_2 + \cdots + r_k = -a_{k-1}/a_k, \\ K_2 &= r_1r_2 + r_1r_3 + \cdots + r_{k-1}r_k = a_{k-2}/a_k, \\ &\dots\dots\dots \\ K_k &= r_1r_2\cdots r_k = (-1)^k a_0/a_k \end{aligned}$$

Proceed now to examine the case $n = k+1$, and

$$\begin{aligned} f(x) &= a_{k+1}x^{k+1} + a_kx^k + \cdots + a_0 \\ &= a_{k+1}(x-r_1)(x-r_2)\cdots(x-r_k)(x-r_{k+1}) \\ &= a_{k+1}(x^k - (r_1 + r_2 + \cdots + r_k)x^{k-1} + (r_1r_2 + r_1r_3 + \cdots + r_{k-1}r_k)x^{k-2} \\ &\quad - \cdots + (-1)^k r_1r_2\cdots r_k)(x-r_{k+1}) \\ &= a_{k+1}(x^{k+1} - (r_1 + r_2 + \cdots + r_{k+1})x^k + (r_1r_2 + r_1r_3 + \cdots + r_k r_{k+1})x^{k-1} \\ &\quad - \cdots + (-1)^{k+1} r_1r_2\cdots r_{k+1}) \\ &= a_{k+1}(x^{k+1} - K'_1x^k + K'_2x^{k-1} - \cdots + (-1)^{k+1} K'_{k+1}) \end{aligned}$$

which implies

$$\begin{aligned} K'_1 &= r_1 + r_2 + \cdots + r_{k+1} = -a_k/a_{k+1}, \\ K'_2 &= r_1r_2 + r_1r_3 + \cdots + r_k r_{k+1} = a_{k-1}/a_{k+1}, \\ &\dots\dots\dots \\ K'_{k+1} &= r_1r_2\cdots r_{k+1} = (-1)^{k+1} a_0/a_{k+1} \end{aligned}$$

Hence, by induction, $f(x) = a_n(x-r_1)(x-r_2)\cdots(x-r_n)$ implies

$$\begin{aligned} S_1 &= r_1 + r_2 + \cdots + r_n = -a_{n-1}/a_n, \\ S_2 &= r_1r_2 + r_1r_3 + \cdots + r_{n-1}r_n = a_{n-2}/a_n, \\ &\dots\dots\dots \\ S_n &= r_1r_2\cdots r_n = (-1)^n a_0/a_n \end{aligned}$$

which completes the proof.

5. Let a, b, c be the three roots of $x^3 + px + q = 0$, and express $(a-b)^2(b-c)^2(c-a)^2$ in terms of p and q .

Solution:

Since, by Th. 5.2.3.4,

$$a + b + c = 0, \quad ab + bc + ca = p, \quad abc = -q$$

it follows that

$$(a-b)^2 = (a+b)^2 - 4ab = (-c)^2 - 4(p - bc - ca) = c^2 - 4p + 4c(a+b) = -4p - 3c^2$$

Likewise $(b-c)^2 = -4p - 3a^2$ and $(c-a)^2 = -4p - 3b^2$.

Hence

$$\begin{aligned} (a-b)^2(b-c)^2(c-a)^2 &= -(4p+3p^2)(4p+3b^2)(4p+3c^2) \\ &= -64p^3 - 48(a^2+b^2+c^2)p^2 - 36(a^2b^2+b^2c^2+c^2a^2)p - 27a^2b^2c^2 \end{aligned}$$

But

$$\begin{aligned} a^2 + b^2 + c^2 &= (a+b+c)^2 - 2(ab+bc+ca) = -2p, \\ a^2b^2 + b^2c^2 + c^2a^2 &= (ab+bc+ca)^2 - 2abc(a+b+c) = p^2, \quad a^2b^2c^2 = q^2 \end{aligned}$$

$$\text{Hence } (a-b)^2(b-c)^2(c-a)^2 = -64p^3 + 96p^3 - 36p^3 - 27q^2 = -(4p^3 + 27q^2)$$

Note. $(a-b)^2(b-c)^2(c-a)^2 = D$ is in fact the discriminant of the given cubic equation (cf. Prob. 34 below).

6. Given the three roots a, b, c of $x^3 - px + q = 0$, construct a cubic equation whose three roots are a^2, b^2, c^2 .

Solution:

Since, by Th. 5.2.3.4, the desired equation must be of the form

$$x^3 - (a^2 + b^2 + c^2)x^2 + (a^2b^2 + b^2c^2 + c^2a^2)x - a^2b^2c^2 = 0$$

and since, from Prob. 5 above,

$$a^2 + b^2 + c^2 = -2p, \quad a^2b^2 + b^2c^2 + c^2a^2 = p^2, \quad a^2b^2c^2 = q^2$$

it follows at once that the equation at issue is

$$x^3 + 2px^2 + p^2x - q^2 = 0$$

7. Solve the following equations simultaneously:

$$(i) \ x + y + z = 9, \quad (ii) \ x^2 + y^2 + z^2 = 41, \quad (iii) \ x^2(y + z) + y^2(z + x) + z^2(x + y) = 180$$

Solution:

From (ii) it follows, using (i),

$$41 = x^2 + y^2 + z^2 = (x + y + z)^2 - 2(xy + yz + zx) = 81 - 2(xy + yz + zx)$$

or

$$xy + yz + zx = 20 \quad (ii')$$

Also it follows from (iii) that, by (i) and (ii),

$$\begin{aligned} 180 &= x^2(y + z) + y^2(z + x) + z^2(x + y) = x^2(9 - x) + y^2(9 - y) + z^2(9 - z) \\ &= 9(x^2 + y^2 + z^2) - (x^3 + y^3 + z^3) = 9 \cdot 41 - (x^3 + y^3 + z^3 - 3xyz + 3xyz) \\ &= 369 - ((x + y + z)(x^2 + y^2 + z^2 - xy - yz - zx) + 3xyz) = 369 - (9(41 - 20) + 3xyz) \end{aligned}$$

or

$$xyz = 0 \quad (iii')$$

Hence, by (i), (ii'), (iii'), and Th. 5.2.3.4, it follows that x, y, z are the three roots of the equation

$$t^3 - 9t^2 + 20t = t(t - 4)(t - 5) = 0$$

which implies six sets of values for x, y, z to satisfy (i), (ii), (iii):

$$(0, 4, 5), \quad (0, 5, 4), \quad (4, 0, 5), \quad (4, 5, 0), \quad (5, 0, 4), \quad (5, 4, 0)$$

8. Find the necessary and sufficient condition that the sum of any two roots of

$$x^4 + px^3 + qx^2 + rx + s = 0 \quad (1)$$

is equal to the sum of the other two roots.

Solution:

Let the four roots of (1) be a, b, c, d ; then, by Th. 5.2.3.4,

$$S_1 = a + b + c + d = -p, \quad (2)$$

$$S_2 = ab + ac + ad + bc + bd + cd = q, \quad (3)$$

$$S_3 = abc + abd + acd + bcd = -r, \quad (4)$$

$$S_4 = abcd = s \quad (5)$$

If $a + b = c + d$ by hypothesis, then, from (2)

$$2(a + b) = 2(c + d) = -p \quad (6)$$

and, from (3) and (4),

$$(a + b)(c + d) + ab + cd = q, \quad ab(c + d) + cd(a + b) = -r$$

where, substituting (6),

$$(p^2/4) + ab + cd = q, \quad (-p/2)(ab + cd) = -r$$

which implies, eliminating $ab + cd$, $(p/2)(q - (p^2/4)) = r$, i.e.,

$$p^3 - 4pq + 8r = 0 \quad (7)$$

which is the desired necessary condition.

Conversely, represent the left-hand side of (7) by $f(p)$, expressing it in terms of a, b, c, d , through (2), (3), (4), and

$$f(p) = -(a+b+c+d)^3 + 4(a+b+c+d)(ab+ac+ad+bc+bd+cd) - 8(abc+abd+acd+bcd)$$

which implies, substituting $a+b = c+d$ by hypothesis,

$$f(p) = -8(a+b)^3 + 8(a+b)((a+b)^2 + (ab+cd)) - 8(a+b)(ab+cd) = 0$$

which implies $f(p)$ has a factor $a+b-c-d$. But, since $f(p)$ is evidently a symmetric polynomial with respect to a, b, c, d , it must contain two other factors $a+c-b-d$ and $a+d-b-c$, which are obtained by interchanging b, c and b, d . Hence

$$f(p) = k(a+b-c-d)(a+c-b-d)(a+d-b-c)$$

where k is a constant, since $f(p)$ is cubic with respect to a, b, c, d . Hence $a+b = c+d$ if (7) holds, i.e. $f(p) = 0$, which implies that (7) is also the sufficient condition at issue, and that (7) is indeed the desired condition.

Note. $a+c = b+d$ or $a+d = b+c$ may be used as hypothesis, yielding exactly the same result, viz. (7).

9. Find the necessary and sufficient condition that the sum of any two roots of the equation

$$x^3 + px + q = 0 \quad (1)$$

be equal to a root of the equation

$$x^2 + rx + s = 0 \quad (2)$$

Solution:

Let the three roots of (1) be a, b, c and the two roots of (2) be d, e ; then consider the following symmetric polynomial S with a factor $(a+b-d)$ of the lowest degree with respect to a, b, c, d, e :

$$S = (a+b-d)(b+c-d)(c+a-d)(a+b-e)(b+c-e)(c+a-e) \quad (3)$$

where $S=0$ if the sum of any two of a, b, c is equal to either d or e , and conversely. Hence the desired condition is obtained if the coefficients of $S=0$ are expressed in terms of p, q, r, s .

Now, by Th. 5.2.3.4,

$$\begin{aligned} (a+b-d)(b+c-d)(c+a-d) &= (a+b)(b+c)(c+a) - ((a+b)(b+c) \\ &\quad + (b+c)(c+a) + (c+a)(a+b))d + 2(a+b+c)d^2 - d^3 \\ &= q - pd - d^3 \end{aligned}$$

and likewise

$$(a+b-d)(b+c-e)(c+a-e) = q - pe - e^3$$

Hence, again by Th. 5.2.3.4,

$$\begin{aligned} S &= (q - pd - d^3)(q - pe - e^3) \\ &= q^2 - pq(d+e) - q(d^3+e^3) + de(p^2 + p(d^2+e^2) + d^2e^2) \\ &= q^2 + pqr + qr(r^2-3s) + s(p^2 + p(r^2-2s) + s^2) \end{aligned}$$

Thus the desired condition is $q^2 + qr(p+r^2-3s) + s(p^2 + pr^2 - 2ps + s^2) = 0$.

10. Solve

$$3x^4 - 20x^3 + 27x^2 + 26x - 24 = 0 \quad (1)$$

where the product of two roots is equal to 2.

Solution:

Since, by Th. 5.2.3.4, the product of all four roots is known to be $-24/3 = -8$, the product of the other two roots of (1) must be $-8/2 = -4$. Hence, using Th. 5.2.3.4 again, let

$$x^4 - (20/3)x^3 + 9x^2 + (26/3)x - 8 = (x^2 + px + 2)(x^2 + qx - 4) \quad (2)$$

from which it follows

$$p + q = -20/3 \quad \text{and} \quad -4p + 2q = 26/3$$

which in turn imply

$$p = -11/3 \quad \text{and} \quad q = -3$$

which change (1) into

$$3(x^2 - (11/3)x + 2)(x^2 - 3x - 4) = 0$$

i.e.

$$(3x - 2)(x - 3)(x - 4)(x + 1) = 0$$

which yields the desired roots: $2/3, 3, 4, -1$.

11. Prove Th. 5.2.3.5.

PROOF:

Let $a < b$ and M be the set of every x which satisfies

$$f(x) < 0, \quad a \leq x < b \tag{1}$$

Then, since $x = a$ satisfies (1), M is not empty, and also, since the interval is bounded above, there must exist a l.u.b. (or sup. (cf. Df. 2.4.1.6)), say u , in M such that

$$a \leq u \leq b \tag{2}$$

Now let the given polynomial $f(x)$ be

$$f(x) = \sum_k d_k x^k, \quad k = 0, 1, \dots, n$$

and assume $f(u) < 0$. Then a small positive number e can be found (cf. §5.1.2, Prob. 45) such that

$$|d_k(u + e')^{n-k} - d_k u^{n-k}| < f(u)/n, \quad k = 0, 1, \dots, n$$

for an arbitrary real number e' such that $|e'| \leq e$. Add these n inequalities, and

$$|f(u + e') - f(u)| < f(u)$$

which, since $f(u) > 0$, implies

$$f(u + e') < 0 \tag{3}$$

On the other hand, since u is a l.u.b. of M , it follows that $u + e \notin M$ for some e' such that $e' < 0$ and $|e'| \leq e$, which implies, when substituted in (1),

$$f(u + e') > 0 \tag{4}$$

which clearly contradicts (3).

Furthermore, assume $f(u) < 0$. Then $u < b$, since $f(b) > 0$ by hypothesis. Now take a sufficiently small positive number e such that $u + e < b$, and

$$|d_k(u + e')^{n-k} - d_k u^{n-k}| < |f(u)|/n, \quad k = 0, 1, \dots, n$$

for an arbitrary real number e' such that $|e'| \leq e$. Add these n inequalities, and

$$|f(u + e') - f(u)| < |f(u)|$$

which, since $f(u) < 0$, implies

$$f(u + e') < 0$$

which, when $e' = e$, in turn implies

$$f(u + e) < 0$$

which is contradictory to the initial assumption that u is a l.u.b. of M . Hence $f(u) = 0$ is the only alternative, which completes the proof with $u = c$.

Note. The assumption $a > b$, instead of $a < b$, brings forth the same result, as can be readily verified. It must be noted, too, that the result is a special case of the Mean-value Theorem of the Calculus and allows a geometric proof, which is rather intuitively obvious; viz. a continuous curve which passes through two points $(a, f(a))$ and $(b, f(b))$ on the opposite sides of the X -axis must cross the same axis at least once and, in general, an odd number of times between $x = a$ and $x = b$.

12. Prove Th. 5.2.3.6.

PROOF:Let n be an odd number and

$$f(x) = \sum_k c_k x^k, \quad k = 0, 1, \dots, n$$

with the stipulation $c_n = 1$, which does not affect the generality of the problem. Since n is odd, a positive number a can be made sufficiently large such that

$$(-a)^n/n + c_k(-a)^{n-k} = -a^n/n + c_k(-a)^{n-k} < 0, \quad k = 0, 1, \dots, n$$

by letting, e.g., $a > (-1)^{n-k} n c_k$. Adding these together, $f(-a) < 0$.

Likewise, a positive number b can be taken large enough to yield

$$b^n/n + c_k b^{n-k} > 0, \quad k = 0, 1, \dots, n$$

by letting, e.g., $b > -n c_k$. Adding these together, $f(b) > 0$.

Hence, by Th. 5.2.3.5, there exists a real root between $-a$ and b , which completes the proof.

13. Every polynomial with real coefficients of even degree has at least one positive root and one negative root if its leading coefficient is positive and its constant term is negative.

PROOF:

Let the given polynomial be

$$f(x) = \sum_k a_{2n-k} x^{2n-k} = x^{2n} + b_{2n-1} x^{2n-1} + \dots + b_0, \quad k = 0, 1, \dots, n$$

where $b_{2n-j} = a_{2n-j}/a_{2n}$, $j = 0, 1, \dots, n-1$. Then $f(+\infty) > 0$ and $f(0) = b_0 < 0$, by hypothesis. Hence, by Prob. 11, there must exist a real root between 0 and $+\infty$, i.e. a positive root.

Likewise, since $f(-\infty) > 0$ and $f(0) < 0$, there must exist a real root between 0 and $-\infty$, i.e. a negative root.

Hence $f(x)$, of even degree, has at least one positive root and one negative root, which completes the proof.

14. Find the necessary and sufficient condition that

$$f(x) \equiv x^2 + 2pqx + p^2 + q^2 - 1 = 0 \quad (1)$$

where $p, q \in \bar{R}$ has two real roots, the absolute values of which are less than 1.

Solution:The necessary and sufficient condition that $f(x) = 0$ has two real roots is

$$D = p^2 q^2 - p^2 - q^2 + 1 = (p^2 - 1)(q^2 - 1) \geq 0 \quad (2)$$

where D is of course the discriminant of $f(x)$ of degree two. Since ± 1 must lie outside the desired interval,

$$f(1) = (p + q)^2 > 0 \quad \text{and} \quad f(-1) = (p - q)^2 > 0$$

which imply that the desired condition must contain

$$p^2 \neq q^2 \quad (3)$$

Also, since the sum of the real roots of $f(x)$ must be less than 2,

$$p^2 q^2 < 1 \quad (4)$$

must be the case, which implies $q^2 < 1$ if $p^2 < 1$ in compliance with (2). Hence the desired condition as a whole is, from (3) and (4),

$$p^2 < 1, \quad q^2 < 1, \quad p^2 \neq q^2$$

15. If m designates the maximal value of all absolute values of the real coefficients of the equation

$$f(x) \equiv x^n + a_{n-1}x^{n-1} + \cdots + a_0 = 0$$

then all roots of $f(x)$ are found between $\pm(m+1)$.

PROOF:

Since, directly computing a series,

$$1/(m+1) + 1/(m+1)^2 + \cdots + 1/(m+1)^n = (1 - 1/(m+1)^n)/m$$

it follows that

$$\begin{aligned} f(m+1) &= (m+1)^n + a_{n-1}(m+1)^{n-1} + \cdots + a_0 \\ &= (m+1)^n (1 + a_{n-1}/(m+1) + \cdots + a_0/(m+1)^n) \\ &\geq (m+1)^n (1 - m/(m+1) - \cdots - m/(m+1)^n) \\ &= (m+1)^n (1 - m(1/(m+1) + \cdots + 1/(m+1)^n)) \\ &= (m+1)^n (1 - (1 - 1/(m+1)^n)) = 1 \end{aligned}$$

i.e. $f(m+1) \geq 1 > 0$, and also $f(+\infty) > 0$. Hence it follows from Th. 5.2.3.5 that $f(x)$ has no root which is greater than $m+1$.

Likewise, taking $+$ for even degree and $-$ for odd degree,

$$\pm f(-x) = x^n - a_{n-1}x^{n-1} + \cdots + (-1)^n a_0$$

and

$$\begin{aligned} \pm f(-(m+1)) &= (m+1)^n - a_{n-1}(m+1)^{n-1} + \cdots + (-1)^n a_0 \\ &\geq (m+1)^n - (m+1)^{n-1} - \cdots - m = 1 \end{aligned}$$

Hence $\pm f(-(m+1)) > 0$, which implies, together with $\pm f(-\infty) > 0$, that $f(x)$ has no root between $-\infty$ and $-(m+1)$. All roots of $f(x)$ must thus lie between $\pm(m+1)$, completing the proof.

16. Any rational root of a monic polynomial (cf. Df. 4.1.2.5.14) with integral coefficients is an integer and an exact divisor of the nonzero constant term of the polynomial.

PROOF:

Let the given polynomial be represented by

$$f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0 = 0 \quad (1)$$

and assume p/q to be a rational root of (1) in its lowest terms. Then

$$f(p/q) = (p/q)^n + a_{n-1}(p/q)^{n-1} + \cdots + a_0 = 0$$

which implies

$$p^n/q = -(a_{n-1}p^{n-1} + a_{n-2}p^{n-2}q + \cdots + a_0q^{n-1}) \quad (2)$$

which in turn implies that the right-hand side of (2) is an integer, since every term is an integer, while the left-hand side of (2) is a fraction in its lowest terms. This contradiction leads to the conclusion that any rational root of (1) must be an integer.

Furthermore, if c is thus an integral root of (1), then

$$c^n + a_{n-1}c^{n-1} + \cdots + a_1c + a_0 = 0$$

which implies

$$a_0/c = -(c^{n-1} + a_{n-1}c^{n-2} + \cdots + a_1) \quad (3)$$

which in turn implies that the left-hand side of (3) is an integer and therefore a_0/c also must be an integer. Hence a_0 is an exact divisor of a_0 .

17. If $f(x)$ is a polynomial with integral coefficients, and if $f(0)$ and $f(1)$ are both odd numbers, then $f(x) = 0$ cannot have integral roots.

PROOF:

Let $f(x) = \sum_k a_k x^k$, $k=0,1,\dots,n$; then, by hypothesis, $f(0) = a_0$ and $f(1) = \sum_k a_k$ are both odd.

Hence $a_{n-1} + \cdots + a_1 + a_0$ must be an even number and there must be an even number of odd numbers among a_{n-1}, \dots, a_1, a_0 .

Furthermore, if $f(x) = 0$ has an integral root, say, r , then, by Prob. 16 above, it must be a factor of a_0 and also an odd number, since a_0 is odd. Hence r^n, r^{n-1}, \dots, r^2 are all odd, which implies that $a_k r^k$ is odd if a_k is odd, and even if a_k is even. This in turn implies that the number of odd numbers among $a_n r^n, a_{n-1} r^{n-1}, \dots, a_1 r$ is equal to the number of odd numbers among a_n, a_{n-1}, \dots, a_1 . But the number of the latter is even, as has already been found out; the number of the former is thus also even. It follows, then, that

$$a_n r^n + a_{n-1} r^{n-1} + \dots + a_1 r$$

is even, and, consequently, that $f(r) \neq 0$, which completes the proof.

18. Prove Th. 5.2.3.7.

PROOF:

Let the given polynomial be

$$f(x) \equiv \sum a_k x^k = 0, \quad k = 0, 1, \dots, n \quad (1)$$

where a_n may not be 1 and $a_0 \neq 0$. If c is an integral root of (1), then

$$a_n c^n + a_{n-1} c^{n-1} + \dots + a_1 c + a_0 = 0$$

which immediately implies, as in Prob. 16 above,

$$a_0/c = -(a_n c^{n-1} + a_{n-1} c^{n-2} + \dots + a_1) \quad (2)$$

which implies that a_0/c must be an integer, since the right-hand member of (2) is an integer. Hence c is an exact divisor of a_0 , which completes the proof.

19. Prove Th. 5.2.3.8.

PROOF:

Let the polynomial in question be

$$f(x) \equiv \sum a_k x^k = 0, \quad k = 0, 1, \dots, n \quad (1)$$

where $a_n > 0$ and $a_0 \neq 0$ may be assumed and also, by hypothesis, $x = p/q$ may be substituted, then multiplied by q^n , resulting in

$$a_n p^n + a_{n-1} p^{n-1} + \dots + a_1 p q^{n-1} + a_0 q^n = 0 \quad (2)$$

which in turn implies, as in Prob. 16, 18,

$$a_0 q^n = -p(a_n p^{n-1} + a_{n-1} p^{n-2} q + \dots + a_1 q^{n-1}) \quad (3)$$

where p is evidently a factor of $a_0 q^n$, hence a factor of a_0 , since p and q , thus p and q^n , have no common factor other than unity, by hypothesis.

Likewise, (2) may be rewritten as

$$a_n p^n = -q(a_{n-1} p^{n-1} + \dots + a_1 p q^{n-2} + a_0 q^{n-1}) \quad (4)$$

yielding a similar conclusion that q is a factor of a_n , which completes the proof.

20. Prove Th. 5.2.3.10.

PROOF:

If $f(x)$ is a polynomial with complex coefficients of degree n , $n \geq 1$, then, by Th. 5.2.3.9, there exists a complex number c_1 such that $f(c_1) = 0$; hence, by Th. 5.2.1.3,

$$f(x) = (x - c_1) g(x) \quad (1)$$

If $g(x)$ of (1) has degree greater than 1, then, taking a similar step,

$$g(x) = (x - c_2) h(x) \quad (2)$$

for some $c_2 \in C$. Combining (1) and (2),

$$f(x) = (x - c_1)(x - c_2) h(x)$$

Hence, if $f(x)$ is of degree n , then, by induction,

$$f(x) = (x - c_1)(x - c_2) \cdots (x - c_n)$$

for exactly n complex numbers $c_1, c_2, \dots, c_n \in C$, which completes the proof.

Note. Any polynomial over C is thus completely reducible over C , and all prime polynomials are of degree 1 (cf. Th. 5.2.1.15-16).

21. Prove Th. 5.2.3.11.

PROOF:

Let $z = a + ib$, where $a, b \in \bar{R}$; then the equation

$$(x - a)^2 + b^2 = 0$$

yields two roots z and $\bar{z} = a - ib$. Hence, by hypothesis and Th. 5.2.1.3, it follows that

$$f(x) = ((x - a)^2 + b^2)g(x) + cx + d$$

where $g(x)$ is a polynomial with real coefficients and $c, d \in \bar{R}$.

Furthermore, since $f(z) = 0$ by hypothesis,

$$c(a + ib) + d = 0$$

which implies

$$ca + d = 0 \quad \text{and} \quad cb = 0$$

which in turn implies

$$c(a - ib) + d = 0$$

Hence $f(a - ib) = f(\bar{z}) = 0$, completing the proof.

Second Proof. Since the mapping $z = a + ib \leftrightarrow \bar{z} = a - ib$ is an automorphism of C (cf. Th. 5.1.3.9) and also $\overline{z_1 + z_2} = \bar{z}_1 + \bar{z}_2$ (cf. §5.1.3, Prob. 13) or, in general, $\overline{z_1 + z_2 + \cdots + z_n} = \bar{z}_1 + \bar{z}_2 + \cdots + \bar{z}_n$, it readily follows that

$$\overline{f(x)} = \overline{\sum_k a_k x^k} = \sum_k \overline{a_k x^k} = \sum_k a_k \overline{x^k} = f(\bar{x}), \quad k = 0, 1, \dots, n$$

which immediately implies that $f(\bar{z}) = 0$ if $f(z) = 0$.

Third Proof. Denote the values of

$$\begin{array}{cccc} a + ib, & (a + ib)^2, & \dots, & (a + ib)^n, \\ u_1 + iv_1, & u_2 + iv_2, & \dots, & u_n + iv_n, \end{array}$$

respectively, where $u_k, v_k \in \bar{R}$, $k = 1, 2, \dots, n$, as well as $a, b \in \bar{R}$. Then

$$\begin{aligned} f(z) &= f(a + ib) \\ &= (a_n u_n + a_{n-1} u_{n-1} + \cdots + a_{n-1} u_1 + a_0) + i(a_n v_n + a_{n-1} v_{n-1} + \cdots + a_1 v_1) \\ &= U + iV \end{aligned}$$

Since it can be readily verified by the binomial theorem that $(a + ib)^k = u_k + iv_k$ implies $(a - ib)^k = u_k - iv_k$, it follows that if

$$f(z) = f(a + ib) = U + iV = 0$$

then $U = 0$ and $V = 0$, by Df. 5.1.3.1a, and consequently,

$$f(\bar{z}) = f(a - ib) = U - iV = 0$$

which completes the proof.

22. Given $f(x) = x^3 - 11x - 20$, find the imaginary roots of $f(x) = 0$.

Solution:

Let $x = a + ib$; then

$$f(x) = (a + ib)^3 - 11(a + ib) - 20 = 0 \tag{1}$$

Simplify (1), and

$$a^3 - 3ab^2 - 11a - 20 + i(3a^2b - b^3 - 11b) = 0$$

which implies, by Df. 5.1.3.1a,

$$a^3 - 3ab^2 - 11a - 20 = 0 \quad (2)$$

and

$$3a^2b - b^3 - 11b = 0 \quad (3)$$

Since $b \neq 0$ by hypothesis, it follows from (3) that

$$b^2 = 3a^2 - 11 \quad (4)$$

which is substituted in (2) to obtain

$$4a^3 - 11a + 10 = 0 \quad (5)$$

where, by Th. 5.2.3.7, a must be an exact divisor of 10. Try the divisors of 10, viz. $\pm 1, \pm 2, \pm 5, \pm 10$, in (5), and $a = -2$ is found to satisfy (5), which in turn implies, by (4), $b = \pm 1$. Hence the desired roots are $-2 + i$ and $-2 - i$, verifying an outcome of Th. 5.2.3.11.

23. If $f(x) = x^4 - 4x^3 + 11x^2 - 14x + 10$ is known to have two roots of the form $a + ib$ and $a + 2ib$, then find all roots of $f(x) = 0$.

Solution:

It follows at once, from Th. 5.2.3.11 and by hypothesis, that $f(x)$ has two other roots of the form $a - ib$ and $a - 2ib$. Hence, by Th. 5.2.3.4,

$$(a + ib) + (a + 2ib) + (a - ib) + (a - 2ib) = 4 \quad (1)$$

and

$$(a + ib)(a + 2ib)(a - ib)(a - 2ib) = 10 \quad (2)$$

From (1) it follows that $a = 1$, which in turn implies through (2),

$$(a^2 + b^2)(a^2 + 4b^2) = (1 + b^2)(1 + 4b^2) = 10$$

which implies $b^2 = 1$ or $b^2 = -9/4$. But, since b is of course real, the former alone is valid, viz.

$$b = \pm 1$$

Hence the four roots of $f(x)$ are: $1 + i, 1 - i, 1 + 2i, 1 - 2i$.

24. Prove Th. 5.2.3.12.

PROOF:

Let all real roots of $f(x)$ be b_1, b_2, \dots, b_r , of the n roots in entirety (some of which may be of some multiplicity). Then, by Th. 5.2.3.11, $n - r = 2s$ is an even number of complex roots of the form $c_1 + id_1, c_1 - id_1, \dots, c_s + id_s, c_s - id_s$. Hence, by Th. 5.1.3.3,

$$\begin{aligned} f(x) &= a_n(x - b_1) \cdots (x - b_r)(x - (c_1 + id_1))(x - (c_1 - id_1)) \cdots (x - (c_s + id_s))(x - (c_s - id_s)) \\ &= a_n(x - b_1) \cdots (x - b_r)((x - c_1)^2 + d_1^2) \cdots ((x - c_s)^2 + d_s^2) \end{aligned}$$

which completes the proof.

25. Prove Th. 5.2.3.13.

PROOF:

If $f(x)$ is factored in R according to Th. 5.2.3.12, then

$$f(x) = a_n(x - b_1) \cdots (x - b_r)((x - c_1)^2 + d_1^2) \cdots ((x - c_s)^2 + d_s^2) \quad (1)$$

where $c_j, d_j \in \bar{R}$. Hence, if real values are substituted in x in (1), the quadratic forms in (1) must be positive and

$$a_n(a - b_1) \cdots (a - b_r), \quad (2)$$

$$a_n(b - b_1) \cdots (b - b_r) \quad (3)$$

have the same sign as $f(a)$ and $f(b)$ respectively. If b_j is between a and b , then it follows from Th. 5.2.3.5 that $a - b_j$ and $b - b_j$ are of different signs; otherwise, they are of the same sign. Hence (1) and (2) are of the same sign if the number of b_j between a and b are even; if not, they are of different signs, yielding the theorem.

26. The equation $f(x) = x^3 - 9x - k(x^2 - 1) = 0$ has three real roots for any value of k .

PROOF:

Since $f(-\infty) > 0$, $f(-1) = 8 > 0$, $f(1) = -8 < 0$, $f(+\infty) > 0$, it follows from Th. 5.2.3.5 that $f(x)$ must have three real roots in the three intervals: $(-\infty, -1)$, $(-1, 1)$, $(1, +\infty)$, which completes the proof.

27. The following polynomial with real coefficients,

$$f(x) = x^n + a_{n-3}x^{n-3} + a_{n-4}x^{n-4} + \cdots + a_0$$

where $a_{n-3} \neq 0$, has at least two complex roots.

PROOF:

Let the n roots of $f(x)$ be r_1, r_2, \dots, r_n ; then, by Th. 5.2.3.3,

$$\sum_i r_i = r_1 + r_2 + \cdots + r_n = 0 \quad (1)$$

$$\sum_{i < j} r_i r_j = r_1 r_2 + r_1 r_3 + \cdots + r_1 r_n + r_2 r_3 + \cdots + r_2 r_n + \cdots + r_{n-1} r_n = 0 \quad (2)$$

Combine (1) and (2) through $(1)^2 - 2 \cdot (2)$, viz.,

$$\left(\sum_i r_i\right)^2 - 2\left(\sum_{i < j} r_i r_j\right) = \sum_i r_i^2 = r_1^2 + r_2^2 + \cdots + r_n^2 = 0 \quad (3)$$

Since $a_{n-3} \neq 0$ by hypothesis, which implies that r_1, r_2, \dots, r_n are not identically zero, it follows that if all roots are real, then $\sum_i r_i^2 > 0$, at once contradicting (3). Hence it cannot be the case that

all roots are real; at least one of r_i , then, must be a complex root. But, then, since $f(x)$ is a polynomial with real coefficients, it follows from Th. 5.2.3.11 that it must contain the conjugate of the first complex root, which makes two complex roots for $f(x)$, completing the proof.

28. If w is the n th imaginary root of unity, then $1 + w + w^2 + \cdots + w^{n-1} = 0$.

PROOF:

Since, by hypothesis, w is an imaginary root of

$$f(x) = x^n - 1 = (x-1)(x^{n-1} + x^{n-2} + \cdots + x + 1) = 0$$

it follows that

$$f(w) = (w-1)(w^{n-1} + w^{n-2} + \cdots + w + 1) = 0$$

which implies $1 + w + w^2 + \cdots + w^{n-1} = 0$, since $w \neq 1$, i.e. $w - 1 \neq 0$, by hypothesis.

29. Given a complex root c of $x^7 - 1 = 0$, find an equation whose roots are $c + c^6$, $c^3 + c^4$, $c^2 + c^5$.

Solution:

Since c is a root of $x^7 - 1 = 0$, viz.,

$$(x-1)(x^6 + x^5 + x^4 + x^3 + x^2 + x + 1) = 0$$

it must be the case that

$$c^7 = 1 \quad \text{and} \quad c^6 + c^5 + c^4 + c^3 + c^2 + c + 1 = 0, \quad (\because c \neq 1)$$

which implies

$$\begin{aligned} (c + c^6) + (c^3 + c^4) + (c^2 + c^5) &= c^6 + c^5 + c^4 + c^3 + c^2 + c = -1, \\ (c + c^6)(c^3 + c^4) + (c^3 + c^4)(c^2 + c^5) + (c^2 + c^5)(c + c^6) \\ &= (c^4 + c^5 + c^9 + c^{10}) + (c^5 + c^6 + c^8 + c^9) + (c^3 + c^6 + c^3 + c^{11}) \\ &= (c^4 + c^5 + c^2 + c^3) + (c^5 + c^6 + c + c^2) + (c^3 + c^6 + c + c^4) \\ &= 2(c^6 + c^5 + c^4 + c^3 + c^2 + c) = -2, \end{aligned}$$

$$\begin{aligned}
(c + c^6)(c^3 + c^4)(c^2 + c^5) &= c^6(c^5 + 1)(c + 1)(c^3 + 1) \\
&= c^6(c^9 + c^8 + c^6 + c^5 + c^4 + c^3 + c + 1) \\
&= c^6(c + c^6 + c^5 + c^4 + c^3 + c^2 + c + 1) = c^6c = 1
\end{aligned}$$

Hence the desired equation is, by Th. 5.2.3.4,

$$x^3 + x^2 - 2x - 1 = 0$$

30. Solve the following binomial equation: $x^n - i = 0$.

Solution:

Since $i = \text{cis}(\pi/2)$ by polar representation, let $x = \text{cis } \theta$; then the given equation is changed to

$$(\text{cis } \theta)^n = \text{cis } (\pi/2)$$

i.e. $\text{cis}(n\theta) = \text{cis}(\pi/2)$, which implies, by Df. 5.1.3.1a,

$$\cos(n\theta) = \cos(\pi/2) \quad \text{and} \quad \sin(n\theta) = \sin(\pi/2)$$

which imply $\theta = (2k\pi + (\pi/2))/n$. Hence the desired solution is

$$x = \text{cis}((2k\pi/n) + (\pi/2n)), \quad k = 0, 1, \dots, n-1$$

31. Solve $x^4 + x^3 + x^2 + x + 1 = 0$.

Solution:

Divide the given equation by x^2 , and

$$(x^2 + 1/x^2) + (x + 1/x) + 1 = 0 \tag{1}$$

Substitute $x + (1/x) = y$ in (1), and

$$(y^2 - 2) + y + 1 = 0, \quad \text{i.e.} \quad y^2 + y - 1 = 0 \tag{2}$$

Solve (2) for y , and $y = (-1 \pm \sqrt{5})/2$, which implies

$$x + 1/x = (-1 + \sqrt{5})/2, \quad \text{i.e.} \quad 2x^2 + (1 - \sqrt{5})x + 2 = 0 \tag{3}$$

and

$$x + 1/x = (-1 - \sqrt{5})/2, \quad \text{i.e.} \quad 2x^2 + (-1 - \sqrt{5})x + 2 = 0 \tag{4}$$

Solve (3) and (4) for x , and the desired roots are found to be

$$((-1 + \sqrt{5}) \pm i \sqrt{10 + 2\sqrt{5}})/4, \quad ((1 + \sqrt{5}) \pm i \sqrt{10 + 2\sqrt{5}})/4$$

Second Solution. Multiply the given equation by $x - 1$, and

$$x^5 - 1 = 0 \tag{1}$$

which implies that the desired roots are to be found among the roots of (1) excluding 1, which are

$$\text{cis}(2\pi/5), \quad \text{cis}(2\pi/5), \quad \text{cis}(4\pi/5), \quad \text{cis}(4\pi/5) \tag{2}$$

Let $\theta = 2\pi/5$; then

$$\sin 2\theta = \sin(2\pi - 3\theta) = -\sin 3\theta$$

$$2 \sin \theta \cos \theta = -\sin \theta \cos 2\theta - \cos \theta \sin 2\theta$$

$$2 \cos \theta = -\cos 2\theta - 2 \cos^2 \theta$$

$$2 \cos \theta = -2 \cos^2 \theta + 1 - 2 \cos 2\theta$$

$$\cos \theta = (-1 + \sqrt{5})/4 \quad (\because \theta \text{ lies in the first quadrant.}) \tag{3}$$

$$\sin \theta = \sqrt{1 - \cos^2 \theta} = \sqrt{10 + 2\sqrt{5}}/4 \tag{4}$$

$$\cos 2\theta = 2 \cos^2 \theta - 1 = -(1 + \sqrt{5})/4 \tag{5}$$

$$\sin 2\theta = \sqrt{1 - \cos^2 2\theta} = \sqrt{10 - 2\sqrt{5}}/4 \tag{6}$$

From (3)-(6) it follows readily that (2) yields the same roots as those in the first solution.

Note. As has already been manifest at the start, the given equation remains unaltered when x is changed into its reciprocal. Such an equation is called a *reciprocal equation* (cf. Supplementary Problems 5.25-26), which is usually divided into two classes according to the coefficients of $f(x) = \sum a_k x^k$, $k = 0, 1, \dots, n$; viz.,

- (i) $a_n = a_0, a_{n-1} = a_1, \dots, a_1 = a_{n-1}, a_0 = a_n$
 (ii) $a_n = -a_0, a_{n-1} = -a_1, \dots, a_1 = -a_{n-1}, a_0 = -a_n$

The above problem belongs to the first type while the problem below belongs to the second.

32. Solve $2x^6 - 11x^5 + 17x^4 - 17x^2 + 11x - 2 = 0$.

Solution:

By Th. 5.2.3.7 and inspection, two roots ± 1 are immediately found to be valid for the given equation. Hence divide the equation by $x^2 - 1$, and

$$2x^4 - 11x^3 + 19x^2 - 11x + 2 = 0 \quad (1)$$

Divide (1) by x^2 , then substitute $y = x + (1/x)$, and

$$2(y^2 - 2) - 11y + 19 = 0, \quad \text{i.e.} \quad 2y^2 - 11y + 15 = 0 \quad (2)$$

Solve (2) for y and obtain $y = 3, 5/2$, which implies

$$x^2 - 3x + 1 = 0 \quad \text{and} \quad 2x^2 - 5x + 2 = 0$$

which yield $x = (3 \pm \sqrt{5})/2, 2, 1/2$.

Hence the desired roots are: $1, -1, 2, 1/2, (3 + \sqrt{5})/2, (3 - \sqrt{5})/2$.

33. Transform an equation

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_0 = 0 \quad (1)$$

into another without the term of degree $n-1$.

Solution:

Substitute $x = y + b$ in (1), and

$$\begin{aligned} f(x) &= a_n(y+b)^n + a_{n-1}(y+b)^{n-1} + \dots + a_0 \\ &= a_n y^n + (na_n b + a_{n-1})y^{n-1} + ((n(n-1)/2!)a_n b^2 + (n-1)a_{n-1}b + a_{n-2})y^{n-2} \\ &\quad + ((n(n-1)(n-2)/3!)a_n b^3 + a_{n-1}b^2 + na_{n-2}b + a_{n-3})y^{n-3} \\ &\quad + \dots + a_n b^n + a_{n-1}b^{n-1} + \dots + a_0 \end{aligned}$$

Hence a substitution $b = -a_{n-1}/na_n$ transforms (1) into

$$f(x) = b_n y^n + b_{n-2} y^{n-2} + \dots + b_0 = 0,$$

where $b_n = a_n$,

$$b_{n-1} = -(n-1)a_{n-1}^2/2na_n + a_{n-2},$$

$$\dots$$

$$b_0 = a_n(-a_{n-1}/na_n)^n + a_{n-1}(-a_{n-1}/na_n)^{n-1} + \dots + a_0$$

34. Prove Th. 5.2.3.18.

PROOF:

Change the given equation

$$a_3 x^3 + a_2 x^2 + a_1 x + a_0 = 0 \quad (1)$$

to be monic, by substituting $a = a_2/a_3, b = a_1/a_3, c = a_0/a_3$; then

$$x^3 + ax^2 + bx + c = 0 \quad (2)$$

which is then transformed, substituting $x = \bar{x} - (a/3)$ (cf. Prob. 33), into

$$\bar{x}^3 + p\bar{x} + q = 0 \quad (3)$$

where $p = -a^3/3 + b$ and $q = 2a^3/27 - ab/3 + c$.

Now let $\bar{x} = y + z$ in (3); then

$$y^3 + z^3 + 3yz(y+z) + p(y+z) + q = 0$$

i.e.

$$y^3 + z^3 + q + (3yz + p)(y+z) = 0 \quad (4)$$

The necessary and sufficient condition that \bar{x} satisfies (3) is, then, that y and z satisfy (4), or what is the same, that y and z satisfy

$$3yz = -p \quad (5)$$

and

$$y^3 + z^3 = -q \quad (6)$$

Since (5) implies $y^3z^3 = -p^3/27$, it follows from this and Th. 5.2.3.4 that the equation

$$t^2 + qt - p^3/27 = 0 \quad (7)$$

has two roots y^3 and z^3 . Hence, from (7),

$$y^3 = -q/2 + \sqrt{r} \quad \text{and} \quad z^3 = -q/2 - \sqrt{r} \quad (8)$$

where $r = q^2/4 + p^3/27$. Substitute (8) in $x = y + z$, and

$$\bar{x} = \sqrt[3]{-q/2 + \sqrt{r}} + \sqrt[3]{-q/2 - \sqrt{r}} = u + v \quad (9)$$

Since the cubic roots of a real number are generally of three different values, the combination of (8) represents actually nine different values, of which, however, only the following three values satisfy (5):

$$\bar{x}_1 = u + v, \quad \bar{x}_2 = uw + vw^2, \quad \bar{x}_3 = uw^2 + vw$$

where $w^3 = 1$ (i.e. $w = (-1 + i\sqrt{3})/2$, $w^2 = (-1 - i\sqrt{3})/2$). Hence $\bar{x}_1, \bar{x}_2, \bar{x}_3$ are the roots of (3), which yield, by reversing the initial substitution $x = \bar{x} - (a/3)$, the desired roots of (1), viz.,

$$x_1 = \bar{x}_1 - (a/3), \quad x_2 = \bar{x}_2 - (a/3), \quad x_3 = \bar{x}_3 - (a/3)$$

35. Solve

$$x^3 + 6x^2 - 3x - 148 = 0 \quad (1)$$

Solution:

Transform the given equation, by a substitution of $x = y - 2$ (cf. Prob. 33), to

$$y^3 - 15y - 126 = 0 \quad (2)$$

Then, by Th. 5.2.3.18, $p = -15$ and $q = 126$, which imply

$$t^2 - 126t + 125 = 0$$

the solutions of which are 1 and 125. Hence $u = 1$ and $v = 5$ ($\because uv = -p/3 = 5$), which in turn imply

$$y_1 = u + v = 6, \quad y_2 = uw + vw^2 = -3 - 2\sqrt{3}i, \quad y_3 = uw^2 + vw = -3 + 2\sqrt{3}i$$

from which it follows, going back to the initial substitution of $x = y - 2$, that the desired roots of (1) are

$$x_1 = 4, \quad x_2 = -5 - 2\sqrt{3}i, \quad x_3 = -5 + 2\sqrt{3}i$$

36. Prove Th. 5.2.3.19.

PROOF:

The given equation is changed first to

$$x^4 + ax^3 + bx^2 + cx + d = 0 \quad (1)$$

by substituting $a = a_3/a_4$, $b = a_2/a_4$, $c = a_1/a_4$, $d = a_0/a_4$, which is further transformed into

$$y^4 + py^2 + qy + r = 0 \quad (2)$$

by substituting $x = y - a/4$ in (1) (cf. Prob. 33). (3)

Let $y = u + v + w$; then

$$\begin{aligned} x^4 &= (x^2)^2 = (u^2 + v^2 + w^2 + 2(uv + vw + wu))^2 \\ &= (u^2 + v^2 + w^2)^2 + 4(u^2 + v^2 + w^2)(uv + vw + wu) + 4(u^2v^2 + v^2w^2 + w^2u^2) + 8uvw(u + v + w) \\ &= 2(u^2 + v^2 + w^2)x^2 + 8uvw x + 4(u^2v^2 + v^2w^2 + w^2u^2) - (u^2 + v^2 + w^2)^2 \end{aligned} \quad (4)$$

since $2(u^2 + v^2 + w^2)x^2 = 2(u^2 + v^2 + w^2)^2 + 4(u^2 + v^2 + w^2)(uv + vw + wu)$.

Since, also from (2),

$$x^4 = -px^2 - qx - r \quad (5)$$

it follows from (4) and (5) that u, v, w can be found to satisfy

$$2(u^2 + v^2 + w^2) = -p, \quad 8uvw = -q, \quad 4(u^2v^2 + v^2w^2 + w^2u^2) - (u^2 + v^2 + w^2)^2 = -r \quad (6)$$

which imply

$$u^2 + v^2 + w^2 = -p/2, \quad u^2v^2 + v^2w^2 + w^2u^2 = p^2/16 - r/4, \quad u^2v^2w^2 = q^2/64 \quad (7)$$

which in turn imply, by Th. 5.2.3.4, that the equation

$$t^3 + (p/2)t^2 + ((p^2/16) - (r/4))t - q^2/64 = 0 \quad (8)$$

has three roots $r_1 = u^2, r_2 = v^2, r_3 = w^2$, i.e.,

$$u = \pm\sqrt{r_1}, \quad v = \pm\sqrt{r_2}, \quad w = \pm\sqrt{r_3} \quad (9)$$

Hence the roots of (2) are those which satisfy (9) and $uvw = -q/8$, viz.,

$$\begin{aligned} y_1 &= \sqrt{r_1} + \sqrt{r_2} + \sqrt{r_3} & y_3 &= -\sqrt{r_1} + \sqrt{r_2} - \sqrt{r_3} \\ y_2 &= \sqrt{r_1} - \sqrt{r_2} - \sqrt{r_3} & y_4 &= -\sqrt{r_1} - \sqrt{r_2} + \sqrt{r_3} \end{aligned} \quad (10)$$

Hence the desired roots of (1) are obtained by (3) and (10), viz. $x_1 = y_1 - a/4$, etc.

Note. The method described above is by Euler while the following method is by Ferrari.

Rewrite (1) as

$$x^4 + ax^3 = -bx^2 - cx - d \quad (2')$$

which is then changed, by adding $a^2x^2/4$ to both sides of (2'), to

$$(x^2 + ax/2)^2 = (a^2/4 - b)x^2 - cx - d \quad (3')$$

If the right-hand side of (3') can be made a perfect square, the desired solution is already at hand.

If not, add $y(x^2 + ax/2) + y^2/4$ to both sides of (3') such that

$$(x^2 + ax/2 + y/2)^2 = (y + a^2/4 - b)x^2 + (ay/2 - c)x + (y^2/4 - d) \quad (4')$$

where the right-hand side will be the square of a linear polynomial $ex + f$ iff

$$(ay/2 - c)^2 - 4(y + a^2/4 - b)(y^2/4 - d) = 0$$

i.e.

$$y^3 - by^2 + (ac - 4d)y + 4bd - a^2d - c^2 = 0 \quad (5')$$

which is called the *resolvent* of (3').

If (5') is solved for y in order to make the right-hand side of (4') equal to $(ex + f)^2$ for properly chosen e and f , (3') is then changed to

$$(x^2 + ax/2 + y/2)^2 = (ex + f)^2$$

which yields two quadratic equations

$$x^2 + ax/2 + y/2 = ex + f \quad \text{and} \quad x^2 + ax/2 + y/2 = -ex - f$$

which finally yield the desired four roots.

37. Solve

$$x^4 + 8x^3 + 22x^2 + 32x + 21 = 0 \quad (1)$$

Solution:

A substitution: $x = y - 2$ transforms (1) to

$$y^4 - 2y^2 + 8y - 3 = 0 \quad (2)$$

which, since $p = -2, q = 8, r = -3$ in this context, yields

$$t^3 - 4t^2 + 16t - 64 = 0 \quad (3)$$

according to the step (8) of Prob. 36. By inspection (through Th. 5.2.3.7), the equation (3) yields the first root 4 (a factor of -64), then the others, $4i$ and $-4i$, which imply, in accordance with

$$\sqrt{t_1}\sqrt{t_2}\sqrt{t_3} = -q = -8$$

that $\sqrt{t_1} = 2, \sqrt{t_2} = \sqrt{2}(1 + i), \sqrt{t_3} = -\sqrt{2}(1 - i)$.

Hence the four roots of (2) are:

$$\begin{aligned} y_1 &= (2 + 2\sqrt{2}i)/2 = 1 + i\sqrt{2} & y_3 &= (-2 + 2\sqrt{2})/2 = -1 + \sqrt{2} \\ y_2 &= (2 - 2\sqrt{2}i)/2 = 1 - i\sqrt{2} & y_4 &= (-2 - 2\sqrt{2})/2 = -1 - \sqrt{2} \end{aligned}$$

which yield, by the initial substitution:

$$x_1 = -1 + i\sqrt{2}, \quad x_2 = -1 - i\sqrt{2}, \quad x_3 = -3 + \sqrt{2}, \quad x_4 = -3 - \sqrt{2}$$

which are the desired roots of (1). (The method adopted here is by Euler.)

Second Solution. Rewrite (1) as

$$(x^2 + 4x + y)^2 = (2y + 16 - 22)x^2 + (8y - 32)x + (y^2 - 21)$$

$$\text{i.e.} \quad (x^2 + 4x + y)^2 = 2(y - 3)x^2 + 8(y - 4)x + (y^2 - 21) \quad (2')$$

where the right-hand side will be a perfect square iff

$$8(y - 4)^2 - (y - 3)(y^2 - 21) = 0$$

$$\text{i.e.} \quad y^3 - 11y^2 + 43y - 65 = 0$$

where, by inspection (and Th. 5.2.3.7), a root 5 is immediately found and the problem is reduced to solve

$$(x^2 + 4x + 5)^2 = 4x^2 + 8x + 4$$

$$\text{i.e.} \quad (x^2 + 4x + 5)^2 = 4(x + 1)^2$$

$$\text{i.e.} \quad x^2 + 4x + 5 = \pm 2(x + 1) \quad (3')$$

Solving $x^2 + 4x + 5 = 2(x + 1)$ and $x^2 + 4x + 5 = -2(x + 1)$, the desired roots are found to be $-1 \pm i\sqrt{2}$ and $-3 \pm \sqrt{2}$ respectively, which are indeed exactly the same as the first result. (This second method is by Ferrari.)

38. Find the necessary and sufficient condition that the second term and the fourth term of the equation

$$x^4 + px^3 + qx^2 + rx + s = 0 \quad (1)$$

vanish under one and the same transformation.

Solution:

Substitute $x = y - a$ in (1), and

$$(y - a)^4 + p(y - a)^3 + q(y - a)^2 + r(y - a) + s = 0 \quad (2)$$

If the coefficients of y^3 and y in (2) are to vanish, then it must be the case that

$$-4a + p = 0, \quad \text{i.e.} \quad a = p/4 \quad (3)$$

$$-4a^3 + 3pa^2 - 2qa + r = 0 \quad (4)$$

Substituting (3) in (4),

$$-4(p/4)^3 + 3p(p/4)^2 - 2q(p/4) + r = 0$$

$$\text{i.e.} \quad p^3 - 4pq + 8r = 0 \quad (5)$$

Conversely, if (5) holds, then the transformation $y = x + (p/4)$ does eliminate the coefficients of x^3 and x . Hence (5) is the desired condition.

39. If a, b, c are the three roots of

$$x^3 + px^2 + qx + r = 0 \quad (1)$$

and if $s_n = a^n + b^n + c^n$, then

$$s_n + ps_{n-1} + qs_{n-2} + rs_{n-3} = 0 \quad (2)$$

where $n \geq 3$.

PROOF:

Multiply (1) by x^{n-3} ($n \geq 3$), and

$$x^n + px^{n-1} + qx^{n-2} + rx^{n-3} = 0$$

which are satisfied by a, b, c , of course. That is,

$$a^n + pa^{n-1} + qa^{n-2} + ra^{n-3} = 0 \quad (3)$$

$$b^n + pb^{n-1} + qb^{n-2} + rb^{n-3} = 0 \quad (4)$$

$$c^n + pc^{n-1} + qc^{n-2} + rc^{n-3} = 0 \quad (5)$$

Add (3), (4), (5), and, by hypothesis,

$$\begin{aligned} (a^n + b^n + c^n) + p(a^{n-1} + b^{n-1} + c^{n-1}) + q(a^{n-2} + b^{n-2} + c^{n-2}) + r(a^{n-3} + b^{n-3} + c^{n-3}) \\ = s^n + ps^{n-1} + qs^{n-2} + rs^{n-3} = 0 \end{aligned}$$

which completes the proof.

40. If $a + b + c = 0$, then $(a^7 + b^7 + c^7)/7 = ((a^2 + b^2 + c^2)/2)((a^5 + b^5 + c^5)/5)$.

PROOF:

The equation whose three roots are a, b, c , as defined by hypothesis, must be of the form, by Th. 5.2.3.4,

$$x^3 + px + r = 0 \quad (1)$$

which implies, again by Th. 5.2.3.4,

$$ab + bc + ca = p \quad (2)$$

$$abc = -q \quad (3)$$

which in turn imply, together with $a + b + c = 0$,

$$a^2 + b^2 + c^2 = (a + b + c)^2 - 2(ab + bc + ca) = -2p \quad (4)$$

Let, as in Prob. 39, $s_n = a^n + b^n + c^n$; then, by (4) and hypothesis,

$$s_1 = 0, \quad s_2 = -2p \quad (5)$$

Furthermore, by (1) (and as in Prob. 39),

$$s_{n+3} = -ps_{n+1} - qs_n, \quad n = 0, 1, \dots \quad (6)$$

where $s_0 = 3$, which implies

$$s_3 = -ps_1 - qs_0 = -3q \quad (7)$$

and, by (5),

$$s_4 = -ps_2 - qs_1 = 2p^2 \quad (8)$$

Hence

$$s_5 = -ps_3 - qs_2 = 3pq + 2pq = 5pq$$

and

$$s_7 = -ps_5 - qs_4 = -5p^2q - 2p^2q = -7p^2q$$

which in turn imply, together with (5),

$$s_7/7 = (-p)(pq) = (s_2/2)(s_5/5)$$

i.e. $(a^7 + b^7 + c^7)/7 = ((a^2 + b^2 + c^2)/2)((a^5 + b^5 + c^5)/5)$, which completes the proof.

Chapter 5.3

*Algebraic Fields

*§5.3.1 Algebraic Extensions

Df. 5.3.1.1 An element of a field F is *algebraic over a field F'* , where $F' \subseteq F$, if it is a root of a nonzero polynomial with coefficients in F' .

Example:

$1/2$, a root of $2x - 1$, is algebraic over the field R of rational numbers; so is $\sqrt{2}$, since it is a root of $x^2 - 2$ (cf. §2.1.1, Prob. 13).

Df. 5.3.1.2 An element of a field F is *transcendental over a field F'* , where $F' \subseteq F$, if it is not algebraic over F' .

Example:

π is transcendental over the field C of complex numbers, since it can be proved, as was done by Lindemann, that it is not to be found as a root of polynomials with coefficients over C ; so is e , as was proved earlier by Hermite. So, again, is any number of the form a^b , where a is neither 0 nor 1 and b is any irrational algebraic number, as was proved by Gelfond.

This dichotomy of the algebraic and the transcendental yields the following theorem and definition which articulate Df. 5.3.1.1.

Th. 5.3.1.3 Any element algebraic over a field F is the root r of one and only one monic polynomial $p(x)$ of degree n , $n \geq 1$, which is irreducible in the integral domain $F[x]$ of all polynomials over F . (Cf. Prob. 1.)

Df. 5.3.1.4 The unique monic polynomial $p(x)$ of Th. 5.3.1.3 is called the *minimal polynomial* of r over F , while r is said to be of degree n over F , sometimes denoted by $n = [r:F]$. (Cf. Prob. 15, 17.)

Example:

$x^2 - 2$ is the minimal polynomial of $\sqrt{2}$ over R while $x^3 - 2x$, $x^4 - 4$, $x^5 - 2x^3$, etc., are not, although $\sqrt{2}$ is one of their roots; $\sqrt{2}$ is thus of degree 2 over R . (Cf. Prob. 3.)

Df. 5.3.1.5 A field, say E , is said to be an *extension* of a field F if F is a subfield of E .

Example:

The field C of complex numbers is an extension of the field \bar{R} of real numbers, which in turn is an extension of the field R of rational numbers, since $\bar{R} \subset C$ and $R \subset \bar{R}$.

Furthermore, by Th. 4.1.2.4.2b, every field is an extension of one of the minimal fields, viz. R and I_p , since each field of characteristic zero may be said to contain R (cf. Th. 4.2.1.4 and Th. 5.1.1.2) just as each field of characteristic p may be said to contain the field I_p (cf. Th. 4.2.1.4).

Df. 5.3.1.6 Given an extension E of a field F and a complex S of F , the intersection of all subfields of E containing both F and S , denoted by $F[S]$ (or $F(S)$), is said to be *generated* by the (field) *adjunction* of S , since it is obtained by adjoining S to F (cf. Df. 4.1.2.5.2, and Prob. 4-6 below).

Since $F(S)$, by definition, is the smallest field containing both F and S , it must be contained in every field which contains both F and S ; hence, evidently,

$$F \subseteq F[S] \subseteq E$$

Example:

The complex number field C is obtained from the real number field \bar{R} by adjoining a set S of one element, viz. an imaginary number $i = \sqrt{-1}$, and as has already been proved (cf. Th. 5.1.3.2), $C = \bar{R}(S)$.

In general, as is obvious from the context, $F[S]$ is also an extension of a field.

Df. 5.3.1.7 If an element a of the set S of Df. 5.3.1.6 is algebraic over a field F , then $F[a]$ is said to be a *simple algebraic extension* of F .

Since a simple extension E of a subfield F is thus generated over F by a single element of E (cf. Th. 5.1.1.2), E can be said to be *algebraic* or *transcendental* over F according as the generating element of E is algebraic or transcendental over F .

Stated otherwise: E is algebraic (or transcendental) over its subfield F if every element in E is algebraic (or transcendental) over F .

Df. 5.3.1.8 If an element b of the set S of Df. 5.3.1.6 is algebraic over $F[a]$ of Df. 5.3.1.7, then $F[a, b]$ is said to be a *multiple algebraic extension* over F , although it is still a simple algebraic extension of $F[a]$.

In general, $F[a_1, a_2, \dots, a_n]$ is a multiple algebraic extension of F if each of the fields

$$F[a_1], \quad F[a_1, a_2], \quad \dots, \quad F[a_1, a_2, \dots, a_n]$$

is a simple algebraic extension of the preceding field, the first of which is F itself. Hence:

Th. 5.3.1.9 If $F_1 = F[a_1]$, $F_2 = F_1[a_2]$, \dots , $F_n = F_{n-1}[a_n]$, where $a_1, a_2, \dots, a_n \in S$ (cf. Df. 5.3.1.6), then $F_n = F[a_1, a_2, \dots, a_n]$. (Cf. Supplementary Problems 5.39-40.)

Example:

$R(\sqrt{3}, \sqrt{5})$, where R is the set of rational numbers, can be constructed by two steps, prescribed by the theorem above, and the elements of this extension will be of the form $a + b\sqrt{3} + c\sqrt{5} + d\sqrt{15}$, where $a, b, c, d \in R$ (cf. Prob. 5-6).

Df. 5.3.1.10 An extension E of a field F is said to be of *finite degree* n over F (cf. Prob. 15) if there exist n elements of E : $\alpha_1, \alpha_2, \dots, \alpha_n$, called a *basis* for E over F , such that every element of E is expressible in the form $\sum c_k \alpha_k$, $k=1, 2, \dots, n$, where $c_k \in F$.

Example:

C , the complex number field, is of degree 2 over the field \bar{R} of real numbers with 1 and i as a basis (cf. Df. 5.3.1.1a and also Df. 5.3.1.11 below).

Df. 5.3.1.11 The elements $\alpha_1, \alpha_2, \dots, \alpha_n$ of an extension E of a field F are said to be *linearly dependent* over F if there exist b_1, b_2, \dots, b_n , not all zero, nor all distinct, in F such that

$$\sum_k b_k \alpha_k = 0, \quad k=1, 2, \dots, n$$

Otherwise, they are said to be *linearly independent* over F (cf. Df. 4.1.3.2.4).

Example:

2 and $\sqrt{3}$ are linearly dependent over \bar{R} , since $2(-\sqrt{3}) + (2)\sqrt{3} = 0$ or $2(\sqrt{3}) + (-2)\sqrt{3} = 0$, where $\pm 2, \pm\sqrt{3} \in \bar{R}$, but they are linearly independent over R since $2b_1 + \sqrt{3}b_2 = 0$ implies $b_1 = b_2 = 0$ over R (cf. Prob. 13).

Th. 5.3.1.12 If each of w_r , $r=1,2,\dots,n+1$, is of the form $\sum_s a_{rs} z_s$, $s=1,2,\dots,n$, where $a_{rs} \in F$ and $z_s \in C$, F and C being an arbitrary field and the complex number field respectively, then the set of w_r is linearly dependent over F . (Cf. Prob. 9.)

This theorem and Df. 5.3.1.10 immediately establish the following theorems.

Th. 5.3.1.13 If an extension E of a field F is of finite degree over F , then the degree is unique. (Cf. Prob. 10.)

Th. 5.3.1.14 If a subset A of an extension E of a field F is linearly dependent on a subset B of n elements of E , then A is linearly dependent on at most n elements of A itself. (Cf. Prob. 11.)

Th. 5.3.1.15 If $c \in C$ is of degree n over a field F , then $F[c]$ in the context of Df. 5.3.1.6 is of degree n over F with $1, c, c^2, \dots, c^{n-1}$ as a basis. (Cf. Prob. 19.)

Example:

$R[\sqrt[5]{3}]$ is of degree 5 over the field R of rational numbers, and accordingly consists of all the numbers of the form $a + b\sqrt[5]{3} + c(\sqrt[5]{3})^2 + d(\sqrt[5]{3})^3 + e(\sqrt[5]{3})^4$, where $a, b, c, d, e \in R$. (Cf. Prob. 23.)

The last theorem is further generalized as follows:

Th. 5.3.2.16 If c_1, c_2, \dots, c_m are of degrees n_1, n_2, \dots, n_m respectively over a field F , then $F[c_1, c_2, \dots, c_m]$ is of degree r over F , where $r \leq n_1 n_2 \cdots n_m$, and a basis for $F[c_1, c_2, \dots, c_m]$ over F is contained among the $n_1 n_2 \cdots n_m$ numbers: $c_1^{k_1} c_2^{k_2} \cdots c_m^{k_m}$, where $k_j, j=1,2,\dots,m$, are nonnegative integers less than $n_1-1, n_2-1, \dots, n_m-1$ respectively. (Cf. Prob. 24.)

These theorems allow a reinterpretation and generalization of Df. 5.2.3.14-15 as follows:

A field F_1 may be extended by successive adjunctions of a root of each of a set of binomial equations

$$x^{n_1} = a_1, \quad x^{n_2} = a_2, \quad \dots, \quad x^{n_r} = a_r$$

where $a_1 \in F_1$,

$$a_2 \in F_2 = F_1[\sqrt[n_1]{a_1}],$$

$$a_3 \in F_3 = F_2[\sqrt[n_2]{a_2}] = F_1[\sqrt[n_1]{a_1}, \sqrt[n_2]{a_2}],$$

$$\dots\dots\dots$$

$$a_r \in F_r = F_1[\sqrt[n_1]{a_1}, \sqrt[n_2]{a_2}, \dots, \sqrt[n_r]{a_r}]$$

The elements of F_r are then expressible in terms of radicals relative to F_1 , since they are equivalent to rational functions of $\sqrt[n_1]{a_1}, \sqrt[n_2]{a_2}, \dots, \sqrt[n_r]{a_r}$ with coefficients in F_1 .

The problem of solving equations by radicals (and rational operations) is thus reduced to the case of solving equations in a field by radicals relative to the field, expressing the roots of the equations in terms of radicals relative to the field at issue, as has already been done with respect to quadratic, cubic, and quartic equations (cf. Th. 5.2.3.17-19).

Solved Problems

1. Prove Th. 5.3.1.3.

PROOF:

By hypothesis and Th. 5.3.1.1, an element, say r , algebraic over F is a root of at least one polynomial $f(x)$ with some (or all) nonzero coefficients over F . If $f(x)$ is reducible, then, by Th. 5.2.1.9, it must be of the form

$$f(x) = a(p_1(x))^{n_1} (p_2(x))^{n_2} \cdots (p_m(x))^{n_m} \quad (1)$$

where a is the leading coefficient of $f(x)$ and $p_k(x)^{n_k}$, $k=1,2,\dots,m$, are monic prime polynomials over F . Since $f(r)=0$, there must exist at least one factor p_j in (1) such that $p_j(r)=0$, which in this context may be represented by the unique monic polynomial $p(x)$ of degree n , where $p(r)=0$.

Suppose that r is a root of any polynomial $f(x)$ with $\deg f(x) < n$. Then, since $f(x)$ is of degree less than that of the irreducible polynomial $p(x)$, it follows that $(f(x), p(x)) = 1$, which in turn implies, by Th. 5.2.1.7,

$$a(x)f(x) + b(x)p(x) = 1 \quad (2)$$

where $a(x), b(x) \in F[x]$. Since $x=r$ implies $f(x)=p(x)=0$, it follows from (2) that $0=1$, which is absurd, of course. Hence r is not a root of any polynomial $f(x)$ of degree less than n ; i.e. $f(r)=0$ is the case only if the degree of $f(x)$ is at least n which is the degree of $p(x)$.

Furthermore, $p(x)$ is unique, since the existence of another $p(x)$, say $p'(x)$, implies that $p(x) - p'(x)$ must be a polynomial over F with r as root and $\deg p'(x) < n$, a contradiction. This completes the proof.

2. If any element algebraic over F is the root of an irreducible monic polynomial in $F[x]$, then it is a root of another polynomial in $F[x]$ iff the latter is a multiple of the former.

PROOF:

Let $p(x)$ be the same monic polynomial obtained by Th. 5.3.1.3 and $g(x)$ be the second polynomial in $F[x]$ with $g(a)=0$, where a is, as r in Prob. 1, algebraic over F . Then, dividing $g(x)$ by $p(x)$,

$$g(x) = q(x)p(x) + r(x), \quad \deg r(x) \leq n-1 \quad (1)$$

But, since $g(a)=p(a)=0$, by hypothesis, it follows that $r(a)=0$ and that $r(x)$ is thus identically zero. Hence

$$g(x) = q(x)p(x) \quad (2)$$

Conversely, every such polynomial as $g(x)$ which has turned out to be a multiple of $p(x)$ must have a as a root, as is obvious in (2).

Note. This proves again the uniqueness of $p(x)$ in Prob. 1, since all proper multiples of $p(x)$ are necessarily reducible in (2).

3. Determine the degree of $r = (1 + \sqrt[3]{3})/2$ over the field R of rational numbers, then find its minimal polynomial over R .

Solution:

By hypothesis, $2r-1 = \sqrt[3]{3}$. Then

$$(2r-1)^3 - 3 = 0 \quad \text{and} \quad 8r^3 - 12r^2 + 6r - 4 = 0$$

Hence r is evidently a root of a polynomial $a(x) = 4x^3 - 6x^2 + 3x - 2$ which is of degree 3 over R .

Furthermore, if $a(x)$ is to be reducible over R , it must have a linear factor with rational coefficient, i.e. a rational root (cf. Th. 5.2.1.4). The only possible rational roots of $a(x)$ are limited, however, by Th. 5.2.3.8; viz. $\pm 2, \pm 1, \pm 1/2, \pm 1/4$. Since none of these is a root of $a(x)$, as can be readily verified, $a(x)$ is thus irreducible over R . Hence the minimal polynomial of r over R is $p(x) = (a(x))/4$.

4. If every element in a set S_1 or a set S_2 is in a set S_3 , and if every element in S_3 is in S_1 or S_2 or both and $F_1 = F[S_1]$ in the context of Df. 5.3.1.6, then $F_1[S_2] = F[S_3]$.

PROOF:

Since $F[S_3]$ contains F and S_3 , by Df. 5.3.1.6, it contains F and S_1 , by hypothesis. Hence $F[S_3]$ must contain F_1 , and since it now contains F_1 and S_2 , it must contain also $F_1[S_2]$.

Conversely, $F_1[S_2]$ contains F , S_1 and S_2 , since it contains F_1 and S_2 . It must thus contain F and S_3 , by hypothesis, and hence $F[S_3]$.

Hence $F_1[S_2] = F[S_3]$.

Note. Df. 5.3.1.9 is in fact a result from this theorem, deduced by induction on n , as can be readily verified (viz. first by adjoining a_1, a_2, \dots, a_{n-1} to F , then adjoining a_n to the result).

5. Given the field R of rational numbers, determine $R[\sqrt{3}, \sqrt{5}]$.

Solution:

In the context of Prob. 4 above, $R[\sqrt{3}] = R_1$ contains R and $\sqrt{3}$, thus all the elements of the form $a + b\sqrt{3}$, where $a, b \in R$, which do constitute a field.

Likewise, $R_1[\sqrt{5}] = R_2$ contains R_1 and $\sqrt{5}$, forming a minimal field that contains R_1 and $\sqrt{5}$, and the elements of R_2 are evidently of the form $c + d\sqrt{5}$, where $c, d \in R_1$.

Hence $R_2 = R[\sqrt{3}, \sqrt{5}]$ consists of all the elements of the form

$$(a_1 + b_1\sqrt{3}) + (a_2 + b_2\sqrt{3})\sqrt{5} = a + b\sqrt{3} + c\sqrt{5} + d\sqrt{15}$$

where $a_1, a_2, b_1, b_2, a, b, c, d \in R$.

6. Determine the algebraic extensions which contain the roots of quadratic and cubic equations.

Solution:

- (i) Given $x^2 + ax + b = 0$, the field of coefficients is $K = R[a, b]$, where R is the rational number field, and its extension $K_1 = K(\sqrt{a^2 - 4b})$ contains the roots of the quadratic equation (cf. Th. 5.2.3.17).
- (ii) Given $x^3 + ax^2 + bx + c = 0$, the field of coefficients is likewise $K = R(a, b, c)$. Since $y = x + (a/3)$ transforms the given equation into $y^3 + py + q = 0$ (cf. Th. 5.2.3.18), where $p, q \in K$, it follows from the same theorem that the roots of the cubic equation are contained in a field extension L , obtained by iterated adjunctions:

$$L = K[\sqrt{r}, u, v, w]$$

$$\text{where } r = (q/2)^2 + (p/3)^3, \quad u = \sqrt[3]{(-q/2) + \sqrt{r}}, \quad v = \sqrt[3]{(-q/2) - \sqrt{r}}, \quad w = \sqrt[3]{-1}.$$

7. If an element α of an extension E of a field F is linearly dependent on $\beta_1, \beta_2, \dots, \beta_n$ of E , but not on $\beta_2, \beta_3, \dots, \beta_n$, then β_1 is linearly dependent on $\alpha, \beta_2, \dots, \beta_n$.

PROOF:

By hypothesis and Df. 5.3.1.11, there must exist $b_k \in F$, $k = 1, 2, \dots, n$, such that

$$\alpha = b_1\beta_1 + b_2\beta_2 + \dots + b_n\beta_n \quad (1)$$

where $b_1 \neq 0$, since $b_1 = 0$ implies that α is linearly dependent on $\beta_2, \beta_3, \dots, \beta_n$, contrary to the hypothesis. Multiply, then, both sides of (1) by b_1^{-1} , and

$$b_1^{-1}\alpha = b_1^{-1}b_1\beta_1 + b_1^{-1}b_2\beta_2 + \dots + b_1^{-1}b_n\beta_n$$

i.e.

$$\beta_1 = b_1^{-1}\alpha + (-b_1^{-1}b_2)\beta_2 + \dots + (-b_1^{-1}b_n)\beta_n$$

which verifies that β_1 is linearly dependent on $\alpha, \beta_2, \dots, \beta_n$, completing the proof.

8. If an element α of an extension E of a field F is linearly dependent on $\beta_1, \beta_2, \dots, \beta_m$ of E , and if each $\beta_j, j=1, 2, \dots, m$, is linearly dependent on $\gamma_1, \gamma_2, \dots, \gamma_n$ of E , then α is linearly dependent on $\gamma_1, \gamma_2, \dots, \gamma_n$.

PROOF:

By hypothesis and Df. 5.3.1.11, there must exist $b_j \in F, j=1, 2, \dots, m$, and $b_{jk} \in F, k=1, 2, \dots, n$, such that

$$(1) \quad \alpha = \sum_{j=1}^m b_j \beta_j \quad \text{and} \quad (2) \quad \beta_j = \sum_{k=1}^n b_{jk} \gamma_k$$

Then, substituting (2) in (1),

$$\alpha = \sum_{j=1}^m b_j \left(\sum_{k=1}^n b_{jk} \gamma_k \right) = \sum_{k=1}^n \left(\sum_{j=1}^m b_j b_{jk} \right) \gamma_k = \sum_{k=1}^n c_k \gamma_k$$

where $c_k = \sum_{j=1}^m b_j b_{jk}$ and evidently $c_k \in F$. Hence the proof is complete.

Note. The linear dependency of the elements of a field extension thus satisfies transitivity. As a matter of fact, the concept is an equivalence relation, since two sets of $\alpha_1, \alpha_2, \dots, \alpha_m$ and $\beta_1, \beta_2, \dots, \beta_n$ are actually said to be *equivalent* if every $\alpha_i, i=1, 2, \dots, m$, is linearly dependent on the β 's and, conversely, every $\beta_j, j=1, 2, \dots, n$, is linearly dependent on the α 's (cf. Prob. 11-12 below).

9. Prove Th. 5.3.1.12.

PROOF:

If $n=1$, then $w_1 = a_{11}z_1$ and $w_2 = a_{21}z_1$, which imply either $a_{21}w_1 - a_{11}w_2 = 0$ for $a_{11} \neq 0$ or $1 \cdot w_1 + 0 \cdot w_2 = 0$ for $a_{11} = 0$. In either case w_1 and w_2 are thus proved to be linearly dependent over F .

Suppose, then, that the theorem holds up to $n=k$. If $n = k+1$, then, by the assumption and hypothesis,

$$w_r = a_{r1}z_1 + a_{r2}z_2 + \dots + a_{r,k+1}z_{k+1}, \quad r = 1, 2, \dots, k+2$$

which yields, by rearrangement,

$$w_r - a_{r,k+1}z_{k+1} = a_{r1}z_1 + a_{r2}z_2 + \dots + a_{rk}z_k \quad (1)$$

Apply the induction hypothesis to the left-hand side of (1), and

$$b_1(w_1 - a_{1,k+1}z_{k+1}) + b_2(w_2 - a_{2,k+1}z_{k+1}) + \dots + b_{k+1}(w_{k+1} - a_{k+1,k+1}z_{k+1}) = 0$$

where b_1, b_2, \dots, b_{k+1} are in F and are not all zero, which immediately implies

$$b_1w_1 + b_2w_2 + \dots + b_{k+1}w_{k+1} = cz_{k+1} \quad (2)$$

where $c = a_{1,k+1}b_1 + a_{2,k+1}b_2 + \dots + a_{k+1,k+1}b_{k+1}$. If $c=0$, then the proof is already completed.

If $c \neq 0$, then divide (2) by c , and

$$d_1w_1 + d_2w_2 + \dots + d_{k+1}w_{k+1} = z_{k+1} \quad (3)$$

where $d_i = b_i/c, i=1, 2, \dots, k+1$, and obviously $d_i \in F$ since $b_i \in F$. Since the d 's are not all zero, take one of them, say d_1 , and let $d_1 \neq 0$. Then, since w_2, w_3, \dots, w_{k+2} may be treated the same way as w_1, w_2, \dots, w_{k+1} , it follows that

$$e_2w_2 + e_3w_3 + \dots + e_{k+2}w_{k+2} = z_{k+1} \quad (4)$$

where $e_j \in F, j=2, 3, \dots, k+2$, and, subtracting (4) from (3),

$$d_1w_1 + (d_2 - e_2)w_2 + \dots + (d_{k+1} - e_{k+1})w_{k+1} - e_{k+2}w_{k+2} = 0$$

or what is the same,

$$f_1w_1 + f_2w_2 + \dots + f_{k+2}w_{k+2} = 0$$

where $f_1 = e_1, f_2 = d_2 - e_2, \dots, f_{k+2} = e_{k+2}$, and not all zero, of course, since $f_1 = e_1 \neq 0$. Since the theorem is thus verified up to $n = k+1$, the proof is complete by induction.

10. Prove Th. 5.3.1.13.

PROOF:

Let E be of degree n over F and also simultaneously of degree $m < n$ over F . Then, by this assumption, E has two bases, say $\alpha_1, \alpha_2, \dots, \alpha_n$ and $\beta_1, \beta_2, \dots, \beta_m$ over F , and, by Th. 5.3.1.10,

$$\alpha_i = c_{i1}\beta_1 + c_{i2}\beta_2 + \dots + c_{im}\beta_m, \quad i = 1, 2, \dots, n$$

where $c_{ij} \in F$, $j = 1, 2, \dots, m$. Since $n \geq m + 1$, it follows from Th. 5.3.1.12 that

$$d_1 \alpha_1 + d_2 \alpha_2 + \dots + d_{m+1} \alpha_{m+1} = 0$$

where $d_k \in F$, $k = 1, 2, \dots, m+1$, and not all zero. Hence

$$d_1 \alpha_1 + d_2 \alpha_2 + \dots + d_{m+1} \alpha_{m+1} + 0 \cdot \alpha_{m+2} + \dots + 0 \cdot \alpha_n = 0$$

while also

$$0 \cdot \alpha_1 + 0 \cdot \alpha_2 + \dots + 0 \cdot \alpha_n = 0$$

This obviously contradicts Df. 5.3.1.10, since the expression of 0 in terms of the basis $\alpha_1, \alpha_2, \dots, \alpha_n$ is given here in two ways and thus no longer unique. Hence the initial assumption has turned out to be untenable, which completes the proof.

11. If a subset A of an extension E of a field F is linearly dependent on a subset B of n elements of E , then A is linearly dependent on at most n elements of A itself.

PROOF:

Let \mathbf{K} be the class of subsets of E which consist of at most n elements of E and are also linearly equivalent (cf. Prob. 8 note above) to B , and C be one of the sets of \mathbf{K} which contain the largest number of elements of A . Then A is linearly dependent on C , since A is linearly dependent on B .

Now let $D = A \cap C$. If A is not linearly dependent on D , then there must exist at least one element α of A which is not linearly dependent on D . But, since α is linearly dependent on C , C' is not contained in D if C' denotes one of the smallest nonzero subsets of C . Let then all the elements of C' be $\beta_1, \beta_2, \dots, \beta_m$, and in particular $\beta_1 \notin D$. Then α is linearly dependent on $\beta_1, \beta_2, \dots, \beta_m$, but it is not, by the definition of C' itself, linearly dependent on $\beta_2, \beta_3, \dots, \beta_m$. Hence, by Prob. 7, β_1 is linearly dependent on $\alpha, \beta_2, \dots, \beta_m$.

Hence, if C'' denotes a set which contains the elements of C except β_1 , which is now replaced by α in C'' , then C'' and C are linearly equivalent, and C'' consists of at most n elements, yet contain one more element α of A than C , which is evidently a contradiction.

Hence A must be linearly dependent on D , and since $D \subseteq C$ and C consists of at most n elements, the proof is complete.

Note. This theorem necessarily defines that a subset A of an extension E of a field F is linearly independent over F if A is not linearly dependent on any of its proper subsets (cf. Prob. 12-13 below).

12. If A is a finite set, then it is linearly dependent on a linearly independent set B in A .

PROOF:

Let B be one of the minimal subsets of A which are linearly dependent on A . Then, if B is linearly dependent on its proper subset C , it follows from Prob. 8 above that A must be linearly dependent on C , which is obviously contradictory to the definition of B . Hence B must be linearly independent.

13. The necessary and sufficient condition that a finite number of elements $\alpha_1, \alpha_2, \dots, \alpha_n$ of an extension E of a field F are linearly independent over F is that

$$a_1 \alpha_1 + a_2 \alpha_2 + \dots + a_n \alpha_n = 0 \quad (1)$$

where $a_k \in F$, $k = 1, 2, \dots, n$, implies $a_1 = a_2 = \dots = a_n = 0$.

PROOF:

If some a_k , say a_1 , is not zero, then

$$\alpha_1 = -a_1^{-1} a_2 \alpha_2 - a_1^{-1} a_3 \alpha_3 - \dots - a_1^{-1} a_n \alpha_n$$

Hence, by Df. 5.3.1.11 and Prob. 7 above, the set of $\alpha_1, \alpha_2, \dots, \alpha_n$ is linearly dependent on its proper subset of $\alpha_2, \alpha_3, \dots, \alpha_n$; it is thus not linearly independent.

Conversely, if the set of $\alpha_1, \alpha_2, \dots, \alpha_n$ is linearly dependent on its proper subset, say, $\alpha_2, \alpha_3, \dots, \alpha_m$, where $m \leq n$, then

$$\alpha_1 = b_2\alpha_2 + b_3\alpha_3 + \dots + b_m\alpha_m$$

where $b_j \in F$ and $2 \leq j \leq m$, which immediately implies

$$1 \cdot \alpha_1 + (-b_2)\alpha_2 + \dots + (-b_m)\alpha_m + 0 \cdot \alpha_{m+1} + \dots + 0 \cdot \alpha_n = 0$$

Hence (1) holds even for $\alpha_i \in F$, not all zero.

Hence $\alpha_1, \alpha_2, \dots, \alpha_n$ are linearly independent iff $\alpha_1 = \alpha_2 = \dots = \alpha_n = 0$.

14. If A and B are two finite sets, each of which is linearly independent, and if A and B are linearly equivalent, then they have the same number of elements.

PROOF:

Let A and B consist of m and n elements respectively, and assume $m > n$. Then A is linearly dependent on B , which implies, by Prob. 11 above, that A is also linearly dependent on a subset A' of A consisting of at most n elements. But, since $m > n$, A' must be a proper subset of A , which contradicts the hypothesis that A is linearly independent.

A similar contradiction results from $m < n$.

Hence it must be the case that $m = n$, i.e. A and B have the same number of elements.

15. If E is a finite extension of a field F , then every element of E is uniquely expressible in terms of a basis over F .

PROOF:

By Prob. 12, E is linearly dependent on a set of $\alpha_1, \alpha_2, \dots, \alpha_n$, which is linearly independent over F and may be considered a basis of E over F . Then, by Th. 5.3.1.10, every element, say α , of E is expressible in terms of $\alpha_1, \alpha_2, \dots, \alpha_n$, viz.,

$$\alpha = a_1\alpha_1 + a_2\alpha_2 + \dots + a_n\alpha_n \quad (1)$$

where $a_k \in F$, $k = 1, 2, \dots, n$.

Now suppose that the expression (1) is not unique such that

$$\alpha = b_1\alpha_1 + b_2\alpha_2 + \dots + b_n\alpha_n \quad (2)$$

where $b_k \in F$, $k = 1, 2, \dots, n$, and in general $a_k \neq b_k$. Then, by (1) and (2), it follows that

$$(a_1 - b_1)\alpha_1 + (a_2 - b_2)\alpha_2 + \dots + (a_n - b_n)\alpha_n = 0$$

where $(a_k - b_k) \in F$, $k = 1, 2, \dots, n$. But, then, it must follow from Prob. 13 that

$$a_1 = b_1, \quad a_2 = b_2, \quad \dots, \quad a_n = b_n$$

and in general, $a_k = b_k$, contradicting itself. Hence the expression (1) must be unique.

Note. Since it follows from Prob. 14-15 that any two bases of a finite extension E of a field F must be linearly equivalent, $n = [E:F]$ is unique, defining E to be uniquely an extension, of degree n , of F .

16. If E is an extension, of degree n , of a field F , then the maximal number of the elements of E which are linearly independent over F is exactly n .

PROOF:

Let $\alpha_1, \alpha_2, \dots, \alpha_n$ be a basis of E over F and $\beta_1, \beta_2, \dots, \beta_m$ be an arbitrary subset of E which is linearly independent over F . Then the latter set must be linearly dependent on the former set, and it follows from Prob. 11 that $m \leq n$.

On the other hand, by hypothesis, there already exists a basis of n elements which are linearly independent over F , viz. $\alpha_1, \alpha_2, \dots, \alpha_n$. Hence it must be the case that n is the maximal number of linearly independent elements of E over F .

Note. Stated otherwise: E of F , of degree n , implies that no $n+1$ elements of E can ever be linearly dependent over F .

17. If E' is a finite extension of a field F and E is a finite extension of E' such that $F \subseteq E' \subseteq E$, then E is a finite extension of F and $[E:F] = [E:E'] [E':F]$.

PROOF:

Let $\alpha_1, \alpha_2, \dots, \alpha_m$ be a basis of E' over F and $\beta_1, \beta_2, \dots, \beta_n$ a basis of E over E' ; then the proof is complete if mn elements

$$\alpha_i \beta_j, \quad 1 \leq i \leq m \quad \text{and} \quad 1 \leq j \leq n$$

are proved to be a basis of E over F .

Since, by hypothesis, any element γ of E may be represented by

$$\gamma = \sum_j a_j \beta_j, \quad j = 1, 2, \dots, n \quad (1)$$

where $a_j \in E'$ such that

$$a_j = \sum_i b_{ij} \alpha_i, \quad i = 1, 2, \dots, m \quad (2)$$

where $b_{ij} \in F$, it follows from (1) and (2) that

$$\gamma = \sum_j \left(\sum_i b_{ij} \alpha_i \right) \beta_j = \sum_{i,j} b_{ij} \cdot \alpha_i \beta_j \quad (3)$$

which implies that E is linearly dependent on the entire set of $\alpha_i \beta_j$ over F .

Furthermore, since $\sum_{i,j} b_{ij} \cdot \alpha_i \beta_j = 0$ in the context of (3) implies $\sum_j \left(\sum_i b_{ij} \alpha_i \right) \beta_j = 0$, it follows

that, since β_j is linearly independent over E' ,

$$\sum_i b_{ij} \alpha_i = 0, \quad 1 \leq j \leq n$$

and that, since α_i is linearly independent over F ,

$$b_{ij} = 0, \quad 1 \leq i \leq m \quad \text{and} \quad 1 \leq j \leq n$$

which implies that the mn elements of $\alpha_i \beta_j$ are linearly independent over F and, by Df. 5.3.1.10, constitute a basis of E over F , where obviously,

$$mn = [E:F] = [E:E'] [E':F] = m \cdot n$$

completing the proof.

Note. The converse of the theorem evidently holds, viz.: if E is a finite extension of F , and if E' is a subfield of E such that $F \subseteq E' \subseteq E$, then E' is a finite extension of F ; so is E of F .

18. Any finite extension of a field is an algebraic extension of the field.

PROOF:

Let E be a finite extension, of degree n , of a field F and α be any element of E ; then it follows from Prob. 16 that the $n+1$ elements of E ,

$$\alpha^n, \alpha^{n-1}, \dots, \alpha, 1$$

are not linearly independent over F . Hence there must exist $a_k \in F$, $k=0,1,\dots,n$, not all zero, such that

$$a_n \alpha^n + a_{n-1} \alpha^{n-1} + \dots + a_1 \alpha + a_0 = 0 \quad (1)$$

Since a_0, a_1, \dots, a_n are not all zero, there must exist at least one nonzero element among a_1, a_2, \dots, a_n ; for, otherwise, a_0 must be zero, which makes all of the a_k zero, contrary to the initial hypothesis. Hence α in (1) must be a root of a polynomial, of degree greater than 1, which implies, by Df. 5.3.1.1, that α is algebraic over F , completing the proof.

19. Prove Th. 5.3.1.15.

PROOF:

$F[\gamma]$, by Df. 5.3.1.5-6, contains both F and γ , and consequently contains every element of the form

$$a_{n-1}\gamma^{n-1} + a_{n-2}\gamma^{n-2} + \cdots + a_1\gamma + a_0 = \sum_k a_k \gamma^k \quad (1)$$

where $a_k \in F$, $k = 0, 1, \dots, n-1$. If the set of all elements of the form (1) is denoted by E , then E is contained in $F[\gamma]$.

It is evident that E contains γ , since $a_1 = 1$ and $a_k = 0$ for $k \neq 1$ in (1) yield γ itself. Also, $a_k = 0$ for $k \neq 0$ in (1) verifies that every member of F is contained in E .

Now let

$$\alpha = \sum_k a_k \gamma^k \quad \text{and} \quad \beta = \sum_k b_k \gamma^k, \quad k = 0, 1, \dots, n-1$$

where $a_k, b_k \in F$, and also let $\alpha = f(\gamma)$ and $\beta = g(\gamma)$, where

$$f(x) = \sum_k a_k x^k \quad \text{and} \quad g(x) = \sum_k b_k x^k, \quad k = 0, 1, \dots, n-1$$

and $h(x)$ be the minimal polynomial of γ over F . Then, by the Division Algorithm (cf. Th. 5.2.1.2),

$$f(x)g(x) = h(x)q(x) + r(x)$$

where $r(x) = \sum_k c_k x^k$ and $c_k \in F$. Substitute $x = \gamma$, and

$$\alpha\beta = f(\gamma)g(\gamma) = h(\gamma)q(\gamma) + r(\gamma) = \sum_k c_k \gamma^k \quad (2)$$

which is obviously of the form (1). Hence $\alpha\beta \in E$.

Since $h(x)$, by hypothesis, is irreducible, constants and $c h(x)$, where $c \in F$, are the only polynomials with coefficients in F which are factors of $h(x)$. If $\beta = g(\gamma) \neq 0$, then $c h(x)$ cannot be a factor of $g(x)$ and the only common factor of $g(x)$ and $h(x)$ with coefficients in F are constants. Thus, $g(x)$ and $h(x)$ being relatively prime over F , it follows from Th. 5.2.1.7 that there exist two polynomials $a(x)$ and $b(x)$ with coefficients in F such that

$$a(x)g(x) + b(x)h(x) = 1 \quad (3)$$

Then $x = \gamma$ implies that $a(\gamma)g(\gamma) = 1$, so that

$$1/\beta = 1/g(\gamma) = a(\gamma)$$

which in turn implies that

$$\alpha/\beta = f(\gamma)a(\gamma) \quad (4)$$

where $f(\gamma)a(\gamma)$ is of the form (1), as has already been proved by (2). Hence $\alpha/\beta \in E$.

Since evidently also $\alpha \pm \beta \in E$ and all rational operations are thus feasible in E , E satisfies F1-11 and is naturally a field. Furthermore, since E contains both F and γ , as has been shown at the start, E must contain $F[\gamma]$. But, by (1), E is also contained in $F[\gamma]$. Hence $E = F[\gamma]$.

Finally, $\alpha = \beta$ implies that $\sum_k a_k \gamma^k = \sum_k b_k \gamma^k$ and that $\sum_k (a_k - b_k) \gamma^k = 0$. This is possible, however, iff all the $(a_k - b_k)$ are zero, since γ is of degree n over F . Hence, by Df. 5.3.1.11, the set of $1, \gamma, \dots, \gamma^{n-1}$, must be linearly independent over F , i.e. a basis of $E = F[\gamma]$ over F , completing the proof.

20. Given the field R of rational numbers, find the general form of the elements of $R[\sqrt[5]{3}]$.**Solution:**

$\sqrt[5]{3}$ is a root of $x^5 - 3$, which is irreducible over F , since it must otherwise have a linear factor with rational coefficients, i.e. a rational factor. Hence $\sqrt[5]{3}$ is of degree 5 over R and, by Th. 5.3.1.15, $R[\sqrt[5]{3}]$ consists of all the numbers of the form: $a + b\sqrt[5]{3} + c(\sqrt[5]{3})^2 + d(\sqrt[5]{3})^3 + e(\sqrt[5]{3})^4$, where $a, b, c, d, e \in R$.

21. If $E = F[a_1, a_2, \dots, a_n]$, where each of a_i , $i = 1, 2, \dots, n$, is algebraic over a field F , then E is a finite extension of F .**PROOF:**

By Th. 5.3.1.15, $F[a_1]$ is a finite extension of F . Since $F[a_1, a_2]$ is obtained by adjoining a_2 to $F[a_1]$, where a_2 is algebraic over F and now also over $F[a_1]$, it follows that $F[a_1, a_2]$ is a finite extension of $F[a_1]$, and that, by Prob. 17, $F[a_1, a_2]$ is a finite extension of F . Hence, by induction, $F[a_1, a_2, \dots, a_n] = E$ is a finite extension of F , completing the proof.

22. If E is an algebraic extension of a field F , then any algebraic extension of E is also an algebraic extension of F .

PROOF:

Let a be an element of an extension of E and

$$f(x) = b_n x^n + b_{n-1} x^{n-1} + \cdots + b_0$$

where $b_k \in E$, $k=1,2,\dots,n$, be a polynomial in E with a as a root. If $F' = F[b_0, b_1, \dots, b_n]$, then, by Prob. 21, F' is a finite extension of F . Also, since a is algebraic over F' , it follows from Th. 5.3.1.15 that $F'[a]$ is a finite extension of F' and consequently, from Prob. 17, that $F'[a]$ is a finite extension of F . Hence, by Prob. 18, a is algebraic over F .

It is thus proved that any element of an extension of E is algebraic over F if it is algebraic over E itself; hence, by Prob. 21, the proof is complete.

23. Prove Th. 5.3.1.16.

PROOF:

Th. 5.3.1.15 has already proved the theorem for the case of $m=1$, viz. where $\gamma = \gamma_1$.

Assume, therefore, the validity of the theorem up to $\gamma_1, \gamma_2, \dots, \gamma_{m-1}$. Then, since $F[\gamma_1, \gamma_2, \dots, \gamma_m] = G[\gamma_m]$, by Th. 5.3.1.9, where $G = F[\gamma_1, \gamma_2, \dots, \gamma_{m-1}]$, it follows by induction that G is of degree r' over F , where $r' \leq n_1 n_2 \cdots n_{m-1}$, and a basis for G over F is contained among $\gamma_1^{k_1} \gamma_2^{k_2} \cdots \gamma_{m-1}^{k_{m-1}}$, where $0 \leq k_1 \leq n_1 - 1$, $0 \leq k_2 \leq n_2 - 1$, \dots , $0 \leq k_{m-1} \leq n_{m-1} - 1$.

Also, as in Th. 5.3.1.15, γ_m is a root of a polynomial, say $f(x)$, of degree n_m with coefficients in F , since it is of degree n_m over F , by hypothesis. Then, since G contains F , the coefficients of $f(x)$ must be in G , which implies that γ_m is of degree n over G , where $n \leq n_m$.

Now, again by Th. 5.3.1.15, $G[\gamma_m]$ is of degree n over G with $1, \gamma_m, \gamma_m^2, \dots, \gamma_m^{n-1}$ as a basis over G and, by Prob. 17, $G[\gamma_m]$ is of degree r' over F with a basis which consists of nr' of the product $\gamma_1^{k_1} \gamma_2^{k_2} \cdots \gamma_{m-1}^{k_{m-1}} \gamma_m^{k_m}$ where $0 \leq k_1 \leq n_1 - 1$, \dots , $0 \leq k_{m-1} \leq n_{m-1} - 1$, $0 \leq k_m \leq n - 1 \leq n_m - 1$. Hence $F[\gamma_1, \gamma_2, \dots, \gamma_m]$ is of degree $r = nr' \leq n_1 n_2 \cdots n_m$ over F with a basis prescribed by the theorem, and the theorem is complete by induction.

*§5.3.2 Algebraic Numbers

Df. 5.3.2.1 A nonempty set F is an *algebraic number field* (or *algebraic field*) if the sum, difference, product and quotient (except by 0) of any two elements of F is in F itself.

An algebraic (number) field, then, is necessarily a field, as has already been abundantly exemplified by the field R of rational numbers, the field \bar{R} of real numbers, and the field C of complex numbers. The relation between \bar{R} and C in particular readily reveals that an algebraic number field F can be extended by adjoining a number α to it, yielding an extension E of F , viz. $E = F[\alpha]$, which consists of all numbers derived from α and F by rational operations; e.g. $C = \bar{R}[i]$.

Since R is a minimal field (cf. Th. 4.2.1.4) and every number field contains R , Df. 5.3.2.1 may be restated as follows:

Df. 5.3.2.1a Any algebraic number field is a finite extension of R . (Cf. Prob. 3.)

An algebraic number field is necessarily a field, but not conversely; a field may not be an algebraic number field, of course. The following definition of algebraic numbers, for instance, yields an example which forms a field, yet not an algebraic number field.

Df. 5.3.2.2 An *algebraic number* a is a complex number which satisfies a nonzero polynomial equation with rational coefficients, viz.,

$$r_n x^n + r_{n-1} x^{n-1} + \cdots + r_1 x + r_0 = 0$$

where $r_k \in R$, $k=0,1,\dots,n$, and not all $r_k=0$.

Df. 5.3.2.2 is manifestly a special case of Df. 5.3.1.1; viz. F and F' of the latter are specified here to be C and R respectively. An algebraic number is thus any element of C which is algebraic over R (and no other field).

As has already been proved (cf. §2.1, Prob. 13-14), the set A of all real algebraic numbers is denumerable while the set \bar{R} of all real numbers is not (cf. Prob. 1 note). That is, not every real number is algebraic; some, in fact nondenumerably infinitely many, numbers in \bar{R} are thus not algebraic, i.e. transcendental, as are exemplified by π , e , etc. (cf. Prob. 4-6).

Th. 5.3.2.3 The set A of all algebraic numbers is a field. (Cf. Prob. 7.)

The set A is further characterized by the following theorems:

Th. 5.3.2.4 The field A of all algebraic numbers is *algebraically complete* (or *closed*, cf. Df. 5.3.2.10a); viz. every nonzero polynomial equation with coefficients in A has a root in A . (Cf. Prob. 8.)

Th. 5.3.2.5 The field A of all algebraic numbers is not an algebraic number field. (Cf. Prob. 9.)

Certain fields of *some* algebraic numbers, however, do form algebraic number fields, as they are invariably finite extensions of R . The following definition leads to an example of such fields.

Df. 5.3.2.6 A *Gaussian number* is any number of the form: $a + bi$, where $a, b \in R$ and $i = \sqrt{-1}$.

The set \bar{G} of all Gaussian numbers is evidently an extension of R , viz. $\bar{G} = R[i]$, and as such is a subfield of $C = \bar{R}[i]$, articulated by the following theorem:

Th. 5.3.2.7 The set \bar{G} of all Gaussian numbers is an algebraic number field. (Cf. Prob. 10.)

\bar{G} as such is characterized also by many definitions and theorems similar to those with respect to C ; e.g.:

Df. 5.3.2.8 The *conjugate*, *norm*, and *trace* of an element $g = a + bi$, $a, b \in R$, of the field \bar{G} of Gaussian numbers are respectively

$$\bar{g} = a - bi, \quad N(g) = g\bar{g}, \quad T(g) = g + \bar{g}$$

Th. 5.3.2.9 If $g, h \in \bar{G}$, then

$$\begin{array}{lll} \text{(i)} \quad \overline{g+h} = \bar{g} + \bar{h} & \text{(iii)} \quad \overline{g \cdot h} = \bar{g} \cdot \bar{h} & \text{(v)} \quad N(gh) = N(g)N(h) \\ \text{(ii)} \quad \overline{g-h} = \bar{g} - \bar{h} & \text{(iv)} \quad \overline{g/h} = \bar{g}/\bar{h} & \text{(vi)} \quad T(g+h) = T(g) + T(h) \end{array}$$

(Cf. Prob. 11.)

Th. 5.3.2.10 If $g \in \bar{G}$, then g and \bar{g} are the roots of the following polynomial, called the *principal polynomial* of g :

$$x^2 - T(g)x + N(g)$$

(Cf. Prob. 12.)

It can be readily proved that quadratic fields, e.g. $R[\sqrt{2}]$, have properties similar to those of \bar{G} (cf. Prob. 12-14), including that they are also algebraic number fields.

The following examples are those of algebraic numbers which do not form fields, let alone algebraic number fields.

Df. 5.3.2.11 A *Gaussian integer* is any number of the form: $a + bi$, where $a, b \in I$ and $i = \sqrt{-1}$.

I denotes here as elsewhere the set of ordinary integers, which in this specific context may be considered *rational integers* (cf. Df. 4.1.2.3.5). It follows immediately that the sum, difference, and product of Gaussian integers are again Gaussian integers. Hence:

Th. 5.3.2.12 The set \bar{G}^* of all Gaussian integers forms an integral domain, denoted by $I[i]$. (Cf. Prob. 15.)

In parallel to ordinary integral domains, $\bar{G}^* = I[i]$ has many theorems similar to those studied in §4.1.2.2, including the unique factorization theorem (cf. Supplementary Problems 5.50-51).

These Gaussian integers in mind, then, the algebraic integers in general are defined as follows:

Df. 5.3.2.13 Any algebraic number a is an *algebraic integer* if every coefficient of the minimal polynomial (cf. Df. 5.3.1.4) of a is a rational integer; \bar{I} denotes the set of all algebraic integers.

By definition, then, an algebraic integer a is to satisfy the equation of the form

$$p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0 = 0$$

where $a_k \in I$, $k = 0, 1, \dots, n-1$, since a minimal polynomial is monic.

An algebraic integer can be readily identified by the following theorem:

Th. 5.3.2.14 A number a is an algebraic integer if it satisfies over R a monic polynomial equation with rational integral coefficients. (Cf. Prob. 20.)

Example:

$\sqrt[3]{2}$ is an algebraic integer, since it is a root of $x^3 - 2 = 0$, while $1/\sqrt[3]{2}$ is not, since it is a root of $x^3 - (1/2) = 0$.

It must be noted that many odd forms are available for algebraic integers, in particular among those in quadratic fields; e.g. $(-1 + \sqrt{5})/2$, since it is a root of $x^2 + x - 1 = 0$. The latter case is generalized by the following theorem:

Th. 5.3.2.15 Algebraic integers in any quadratic field $R[\sqrt{m}]$, where $m \neq 1$ is an integer free from square factors, are classified into two types:

- (i) If $m \not\equiv 1 \pmod{4}$, then they are of the form $a + b\sqrt{m}$, where $a, b \in I$, and
- (ii) if $m \equiv 1 \pmod{4}$, then they are of the form $(c + d\sqrt{m})/2$, where $c, d \in I$ and $c \equiv d \pmod{2}$. (Cf. Prob. 24.)

Although the set \bar{I} of all algebraic integers is more restricted in comparison with the set \bar{G}^* of all Gaussian integers, \bar{I} and \bar{G}^* share at least one important property in common; viz. they are both integral domains (cf. Th. 5.3.2.12 and Supplementary Problem 5.52).

- *4. If u is a real algebraic number of degree $n > 1$ over R , then there exists a positive number v such that, for any $a, b \in I$ and $b > 0$,

$$|u - a/b| \geq v/b^n$$

PROOF:

If $f(x)$ is the primitive polynomial (cf. Df. 5.2.1.10) of lowest degree satisfied by u , then $f(x)$ must be also of degree n , since it cannot be different by any more than a multiplicative constant from the minimal polynomial for u .

Let w be the maximum of $|f'(x)|$ in the interval $|u - x| \leq 1$ and v be the smaller of 1 and $1/w$. Then it follows at once that $|u - a/b| \geq 1$ implies

$$|u - a/b| \geq v \geq v/b^n$$

which is to be proved.

If, on the other hand, $0 \leq |u - a/b| < 1$, then, by the mean-value theorem,

$$\begin{aligned} |f(u) - f(a/b)| &= |u - a/b| |f'(c)| \\ &\leq |u - a/b| \cdot w \end{aligned}$$

where c lies between u and a/b , i.e. $|u - c| \leq 1$. Now

$$|f(a/b)| \leq |u - a/b| \cdot w$$

since $f(u) = 0$ by hypothesis, and also $f(a/b) \neq 0$ since $f(x)$ is irreducible over R . Furthermore, $|f(a/b)| = d/b^n$, where $d \in I$, since $f(x)$, by hypothesis, is of degree n and has rational coefficients. Hence $n \geq 1$, which implies

$$1/b^n \leq |f(a/b)| \leq w \cdot |u - a/b|$$

and finally,

$$|u - a/b| \geq (1/w)(1/b^n) \geq v/b^n$$

which completes the proof.

- *5. There exist transcendental numbers.

PROOF:

Let an infinite series be

$$s = \sum_r (-1)^r / 2^{r!}, \quad r = 1, 2, \dots, n, \dots$$

and

$$s_k = a_k/b_k = a_k/2^{k!}$$

which denotes the sum of the first k terms of s . Then

$$\begin{aligned} |s - a_k/b_k| &= 1/2^{(k+1)!} - 1/2^{(k+2)!} + \dots \\ &< 1/2^{(k+1)!} \\ &< 1/2^{k \cdot k!} = 1/b_k^k \end{aligned} \tag{1}$$

If s were algebraic over R , of degree $n > 1$, then by (1) and Prob. 4 above,

$$L = b^n \cdot |s - a_k/b_k| \leq b_k^{n-k} \tag{2}$$

where $\lim_{k \rightarrow \infty} L = 0$, which is a contradiction, however. For, by Prob. 4, there exists a number $v > 0$ such that

$$|s - a_k/b_k| \geq v/b_k^n$$

which implies

$$b^n \cdot |s - a_k/b_k| \geq v > 0$$

contradicting (2). Hence s cannot be an algebraic number of degree $n > 1$.

Nor can s be an algebraic number of degree 1, i.e. a rational number. For, if $s = c/d$, where $c, d \in I$ and $d > 0$, then the choice of an odd k such that $2^{k \cdot k!} > d$ implies that the number

$$t = 2^{k!} s d - 2^{k!} d \left(\sum_{r=1}^k (-1)^r / 2^{r!} \right) = 2^{k!} d \left(\sum_{r=k+1}^{\infty} (-1)^r / 2^{r!} \right) \tag{3}$$

is a positive integer while

$$t < 2^{k!} d/2^{(k+1)!} = d/2^{k \cdot k!} < 1 \quad (4)$$

by the choice of k , yielding the contradiction between (3) and (4).

Since s can be neither an algebraic number of degree $n > 1$ nor that of degree 1, it cannot be an algebraic number at all; it must be a transcendental number.

Note. The transcendental numbers of the form s , constructed as above, are due to Liouville. This proof has an obvious advantage over Cantor's existence proof (cf. §2.1, Prob. 13-14), since the latter does not offer an explicit example of transcendental numbers while the former does.

6. If s , constructed as in Prob. 5, is transcendental, so is $s + is$, where $i = \sqrt{-1}$.

PROOF:

Suppose that $s + is$ is not transcendental, i.e. algebraic. Then it must be a root of a polynomial with real coefficients, which is then satisfied by its conjugate: $s - is$ (cf. Th. 5.2.3.11). Hence $s - is$ is also algebraic.

Then, since their sum must be also algebraic (cf. Th. 5.3.2.3),

$$(s + is) + (s - is) = 2s$$

or what is the same, s itself, must be algebraic, which immediately contradicts the hypothesis. Hence $s + is$ must be transcendental.

7. Prove Th. 5.3.2.3.

PROOF:

Let all elements of the set A be algebraic over a field F . The proof is then to show that rational operation is feasible with respect to these elements over F , i.e. the sum, difference, product, and quotient (for nonzero denominator) of any two elements of A are algebraic over F , satisfying any polynomial over F .

If $a, b \in A$, $b \neq 0$, and if $a(x)$ and $b(x)$ are the minimal polynomials over F for a and b respectively, then it follows that $p(x)$ and $q(x)$, defined by Prob. 15 of §5.2.2, are algebraic over F in this context. In the same sense, equate a_1 and b_1 there with a and b here. Then, since $a(x)$ and $b(x)$ are satisfied by $a_1 + b_1$ and $a_1 b_1$ there, they are satisfied by $a + b$ and ab here.

Furthermore, since $-b$ satisfies $b(-x)$, $-b$ is algebraic, and so is $a + (-b) = a - b$. Also, if n is the degree of $b(x)$, then $x^n b(1/x)$ is satisfied by $1/b$, and $1/b$ is algebraic. So, then, is its product: $a(1/b) = a/b$.

Hence $a + b$, $a - b$, ab , a/b are all algebraic over F , and the set A is a field.

8. Prove Th. 5.3.2.4.

PROOF:

Let the coefficients a_k , $k = 0, 1, \dots, n$, of a polynomial equation

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_0 = 0 \quad (1)$$

belong to the field A of algebraic numbers. Then, by Prob. 21-22 of §5.2.1 and Th. 5.3.1.16, an extension $E = R[a_0, a_1, \dots, a_n]$, generated by the coefficients a_k , is a finite extension of the minimal field R of rational numbers.

Since any complex root c of (1) is algebraic over E , by Th. 5.2.3.9, $E[c]$ is then a finite extension of E and, in consequence, of R . Hence, by Prob. 17 of §5.3.1, the element c of E must be algebraic over R , which implies that $c \in A$, proving that A is algebraically closed.

9. Prove Th. 5.3.2.5.

PROOF:

If the field A were an algebraic number field, i.e. a finite (thus simple) extension of R , then let it be of degree over R . But, for instance, a polynomial

$$x^{n+1} - 2$$

which is easily verified by Eisenstein's criterion (cf. Th. 5.2.1.12) to be irreducible over R , readily yields an algebraic number $2^{1/(n+1)}$ of degree $n+1$. The existence of such algebraic numbers contradicts the initial hypothesis through Prob. 17 of §5.3.1. Hence the field A is not an algebraic number field.

10. Prove Th. 5.3.2.7.

PROOF:

Let $g, h \in \bar{G}$ and $g = a + bi$, $h = c + di \neq 0$, where $a, b, c, d \in R$. Then,

- (i) $g + h = (a + bi) + (c + di) = (a + c) + (b + d)i$, where $a + c, b + d \in R$, implying that $g + h \in \bar{G}$.
- (ii) Likewise, $g - h \in \bar{G}$.
- (iii) $g \cdot h = (a + bi)(c + di) = (ac - bd) + (ad + bc)i$, where $ac - bd, ad + bc \in \bar{G}$, implying that $g \cdot h \in \bar{G}$.
- (iv) $g/h = (a + bi)/(c + di) = ((ac + bd)/(c^2 + d^2)) + ((ad - bc)/(c^2 + d^2))i$, where $c^2 + d^2 \neq 0$ and $(ac + bd)/(c^2 + d^2), (ad - bc)/(c^2 + d^2) \in G$, implying that $g/h \in \bar{G}$.

Since \bar{G} is manifestly non-vacuous, and since the rational operations (i)-(iv) on the elements of \bar{G} are closed with respect to \bar{G} , \bar{G} is, by Df. 5.3.2.1, an algebraic number field, completing the proof.

11. Prove Th. 5.3.2.9.

PROOF:

Let $g = a + bi$ and $h = c + di$, as above, where $a, b, c, d \in R$. Then,

- (i) $\bar{g} = a - bi$, $\bar{h} = c - di$, by Df. 5.3.2.8, and

$$\begin{aligned}\overline{g + h} &= \overline{(a + bi) + (c + di)} = \overline{(a + c) + (b + d)i} \\ &= (a + b) - (b + d)i = (a - bi) + (c - di) = \bar{g} + \bar{h}\end{aligned}$$

(ii)-(iv) are proved likewise, (cf. §5.1.3, Prob. 13).

- (v) By Df. 5.2.3.8, $N(g) = g\bar{g}$ and $N(h) = h\bar{h}$; also, by (iii) above (i.e. $\overline{g \cdot h} = \bar{g} \cdot \bar{h}$), it follows that

$$N(gh) = (gh)(\overline{gh}) = g \cdot h \cdot \bar{g} \cdot \bar{h} = g \cdot \bar{g} \cdot h \cdot \bar{h} = N(g)N(h)$$

- (vi) Since $g + h = (a + c) + (b + d)i$, by Prob. 10, (i) above, it follows from Df. 5.3.2.8 that

$$\begin{aligned}T(g + h) &= ((a + c) + (b + d)) + \overline{((a + c) + (b + d)i)} \\ &= (a + c) + (b + d)i + (a + c) - (b + d)i = 2(a + c)\end{aligned}$$

while

$$\begin{aligned}T(g) + T(h) &= ((a + bi) + \overline{(a + bi)}) + ((c + di) + \overline{(c + di)}) \\ &= (a + bi + a - bi) + (c + di + c - di) = 2a + 2c = 2(a + c)\end{aligned}$$

Hence $T(gh) = T(g) + T(h)$.

12. Prove Th. 5.3.2.10.

PROOF:

Let $g = a + bi$; then, by Df. 5.3.2.8,

$$f(x) = x^2 - T(g)x + N(g) = x^2 - (g + \bar{g})x + g\bar{g} = (x - g)(x - \bar{g})$$

Hence $f(x) = 0$ has two roots g and \bar{g} , completing the proof.

Note. $x^2 - 4x + 5$, for instance, is the principal polynomial of $g = 2 + i$, since $\bar{g} = 2 - i$, $T(g) = 4$, $N(g) = 5$.

13(a). The quadratic field $R[\sqrt{m}]$, where $\sqrt{m} \notin R$, represents an algebraic number field.

PROOF:

Let $x, y \in R[\sqrt{m}]$, where $x = a + b\sqrt{m}$, $y = c + d\sqrt{m}$. Then,

- (i)-(ii) $x \pm y = (a + b\sqrt{m}) \pm (c + d\sqrt{m}) = (a \pm c) + (b \pm d)\sqrt{m}$, which implies $x \pm y \in R[\sqrt{m}]$.
- (iii) $x \cdot y = (a + b\sqrt{m})(c + d\sqrt{m}) = (ac + bdm) + (ad + bc)\sqrt{m}$, implying $x \cdot y \in R[\sqrt{m}]$.
- (iv) Since $c - d\sqrt{m} \neq 0$ if $c + d\sqrt{m} \neq 0$,

$$\begin{aligned}x/y &= (a + b\sqrt{m})/(c + d\sqrt{m}) = (a + b\sqrt{m})(c - d\sqrt{m})/((c + d\sqrt{m})(c - d\sqrt{m})) \\ &= ((ac - bdm)/(c^2 - md^2)) + ((ad - bc)/(c^2 - md^2))\sqrt{m}\end{aligned}$$

which proves that $x/y \in R[\sqrt{m}]$.

Hence, by Df. 5.3.2.1, $R[\sqrt{m}]$ forms an algebraic number field.

Note. Some specific subfields of $R[\sqrt{m}]$ are, e.g. $R[\sqrt{2}]$, $R[\sqrt{3}]$, $R[\sqrt{5}]$, etc., which are sometimes called the *real* quadratic fields of $R[\sqrt{m}]$ in general, in contrast to the *imaginary* quadratic fields such as $R[i]$, $R[\sqrt{-2}]$, etc. The algebraic number field \bar{G} of all Gaussian numbers, then, is an imaginary quadratic field.

13(b). If $r = k^2s$, where $k, r, s \in R$ and $k > 0$, then $R[\sqrt{r}] = R[\sqrt{s}]$.

PROOF:

By hypothesis $\sqrt{r} = k\sqrt{s}$, which implies, for any $a, b \in R$,

$$a + b\sqrt{r} = a + bk\sqrt{s}$$

Hence $x \in R[\sqrt{r}]$ implies $x \in R[\sqrt{s}]$, i.e. $R[\sqrt{r}] \subseteq R[\sqrt{s}]$, and $x \in R[\sqrt{s}]$ implies $x \in R[\sqrt{r}]$, i.e. $R[\sqrt{s}] \subseteq R[\sqrt{r}]$. That is

$$R[\sqrt{r}] = R[\sqrt{s}]$$

Note. $R[\sqrt{40}] = R[\sqrt{10}]$, for instance, since $40 = 2^2 \cdot 10$. The two fields, therefore, contain the same elements.

As can be readily verified, the conjugate, norm, and trace of each element of any quadratic field are defined in the same way as Df. 5.3.2.8 with respect to Gaussian numbers, even up to the principal polynomial.

14. If $x, y \in I[i]$, so are $x + y, x - y, xy \in I[i]$.

PROOF:

(i) Let $x = a + bi$, $y = c + di$, where $a, b \in I$; then

$$x + y = (a + bi) + (c + di) = (a + c) + (b + d)i, \quad \text{where } a + c, b + d \in I$$

which implies $x + y \in I[i]$.

(ii)-(iii) are proved likewise (cf. Prob. 10 above).

15. Prove Th. 5.3.2.12.

PROOF:

Prob. 14 above has already assured D1-8 (cf. Df. 4.1.2.2.1) for $I[i]$. As for the rest: for every $x, y, z \in I[i]$, where $x = a + bi$, $y = c + di$, $z = e + fi \neq 0$, and $a, b, c, d, e, f \in I$,

$$\text{D9. } xy = (a + bi)(c + di) = (ac - bd) + (ad + bc)i = (c + di)(a + bi) = yx$$

$$\text{D10. } 1 \in I[i], \text{ where } 1 = 1 + 0 \cdot i.$$

D11. Since $z = e + fi \neq 0$ implies $e \neq 0$ and $f \neq 0$, $xz = yz$ implies $x = y$. For, if $xz = yz$, i.e. $(ae - bf) + (af + be)i = (ce - df) + (cf + de)i$, then $ae - bf = ce - df$ and $af + be = cf + de$, i.e. $(a - c)e - (b - d)f = 0$ and $(a - c)f + (b - d)e = 0$, which implies $a = c$ and $b = d$ in either case, since $e \neq 0$ and $f \neq 0$. Hence $a + bi = c + di$, i.e. $x = y$, if $xz = yz$ and $z \neq 0$.

Hence $I[i]$ is an integral domain.

16. If $x = a + bi$ is an element of $I[i]$, i.e. $a, b \in I$, and $N(x)$ is defined to be $x\bar{x}$ as in Df. 5.3.2.8, then

(i) $N(x) = x^2$ if $x \in I$ as well as $x \in I[i]$.

(ii) $N(xy) = N(x)N(y)$, if $y \in I[i]$.

(iii) $N(x) = 1$ iff x is a unit.

PROOF:

(i) Since $x \in I$, it must be the case that $b = 0$, i.e. $\bar{x} = a$ as well as $x = a$. Hence $N(x) = x\bar{x} = a^2 = x^2$.

(ii) The proof can be carried out exactly as in Prob. 11, (v).

(iii) If x is a unit, then $x \mid 1$ such that $xy = 1$ for some $y \in I[i]$, which implies, by (ii) above, $N(x)N(y) = N(1) = 1$, i.e. $N(x) \mid 1$. Since $N(x)$ must be a non-negative integer, it must be the case that $N(x) = 1$.

Conversely, if $N(x) = 1$, it follows that $a^2 + b^2 = 1$ (cf. Df. 5.3.2.8), which implies, since $a, b \in I$, that either $a = 0$ or $b = 0$. In either case, however, $x = \pm 1, \pm i$, which are evidently units, completing the proof.

17. If $N(x)$, where $x \in I[x]$, is a prime in I , so is x in $I[i]$, but not conversely.

PROOF:

Assume $x = yz$, where $y, z \in I[i]$. Then, by hypothesis and Prob. 16, (ii), $N(x) = N(y)N(z)$ must be prime in I , which implies that either $N(y) = 1$ or $N(z) = 1$, which in turn implies, by Prob. 16, (iii), either y or z is a unit, proving that x must be a prime in $I[i]$.

The converse, however, does not hold. For instance, if 3 is a prime in $I[i]$, it does not follow that $N(3) = 3 \cdot 3 = 9$ is a prime in I . And 3 is indeed a prime in $I[i]$, for the following reason: $3 = xy$, where $x, y \in I[i]$, implies $9 = N(x)N(y)$, while $N(x) \neq 1$ and $N(y) \neq 1$ imply that $N(x) = N(y) = 3$, which in turn implies that either $a^2 + b^2 = 3$ or $c^2 + d^2 = 3$ for $x = a + bi$ and $y = c + di$. In either case, since $a, b, c, d \in I$ and none of them can possibly satisfy the identities, the initial assumption that $N(x) \neq 1$ and $N(y) \neq 1$ is thus proved to be an absurdity. Hence it must be the case that either $N(x) = 1$ or $N(y) = 1$, which immediately implies that 3 is a prime in $I[i]$, although $N(3) \in I$ is not.

18. If $p, q \in I[i]$ and $q \neq 0$, then there exist $r, s \in I[i]$ such that

$$p = sq + r, \quad \text{where } N(r) < N(q)$$

PROOF:

Let $p/q = a + bi$, where $a, b \in R$, and select $m, n \in I$ such that m and n are the rational integers nearest to a and b respectively, i.e. $|a - m| \leq 1/2$ and $|b - n| \leq 1/2$.

Now let $s = m + ni$; then $s \in I[i]$, and

$$(p/q) - s = (a - m) + (b - n)i$$

which implies

$$\begin{aligned} N(p - qs) &= N(q)N((p/q) - s) \\ &= N(q)((a - m)^2 + (b - n)^2) \\ &\leq N(q)((1/4) + (1/4)) < N(q) \end{aligned}$$

i.e. $N(p - qs) < N(q)$. Hence, letting $p - qs = r$, the proof is complete.

Note. The division in $I[i]$, which is an integral domain and fails to be a field, is not unique; i.e. the quotient s and the remainder r may not be unique if $s, r \in I[i]$, as in Prob. 19 below.

19. Divide $2 - i$ by $1 + i$, and give a geometric interpretation.

Solution:

$$(2 - i)/(1 + i) = ((2 - i)(1 - i))/((1 + i)(1 - i)) = (1 - 3i)/2 = (1/2) + (-3/2)i$$

Take 1 and -1 as rational integers nearest to $1/2$ and $-3/2$ respectively; then, in the context of Prob. 18,

$$s = 1 - i \quad \text{and} \quad r = (2 - i) - (1 + i)(1 - i) = -i$$

where $N(r) = 1 < N(1 + i) = 2$.

On the other hand, if 0 and -2 are taken as rational integers nearest to $1/2$ and $-3/2$ respectively, then

$$s = -2i \quad \text{and} \quad r = (2 - i) - (1 + i)(-2i) = i$$

where $N(r) = 1 < N(1 + i) = 2$.

As a matter of fact, there are two more alternatives for the pair of r and s , as can be readily observed in the following geometric representation:

Since the Gaussian integer s is defined by $N(p - qs) < N(q)$ as in Prob. 18, i.e. $N((p/q) - s) < 1$, or what is the same, $|(p/q) - s| < 1$, it follows that, in the complex plane, s is represented by the points whose distances from the point for p/q is less than 1, viz. $-i$, $-2i$, $1 - 2i$, $1 - i$, in this context.

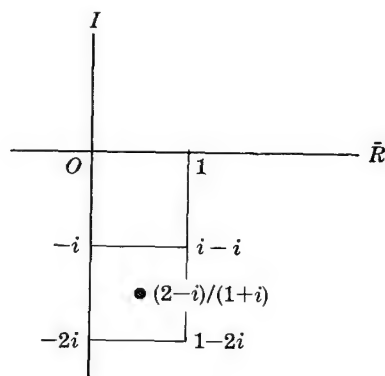


Fig. 5.3.2a

20. If a is algebraic over a field F , then it has a unique minimal polynomial.

PROOF:

If $q(x)$ is a minimal polynomial and $f(x)$ any other polynomial over F satisfied by a , then, by Th. 5.2.1.2, there exist two other polynomials $g(x)$ and $r(x)$ over F such that

$$f(x) = g(x)q(x) + r(x) \quad (1)$$

where $0 \leq \deg r(x) < \deg q(x)$.

Let $x=a$ in (1). Then, since $q(a)=f(a)=0$ by hypothesis, it follows that $r(a)=0$, and that $r(x)=0$; for, otherwise, $q(x)$ cannot be minimal, contrary to the initial hypothesis. Hence $q(x) \mid f(x)$, and a has a minimal polynomial.

Furthermore, the minimal polynomial $q(x)$ is unique. For, if $s(x)$ were any other minimal polynomial of a over F , it must follow likewise that $s(x) \mid f(x)$, which in turn implies that $q(x) = \pm s(x)$. But, then, since $q(x)$ and $s(x)$ are both monic by hypothesis, it must be the case that $q(x) = s(x)$, which completes the proof.

Note. This theorem has an alternative form, or a corollary: $q(x)$, the minimal polynomial of a which is algebraic over F , is contained as a factor in any polynomial satisfied by a over F .

21. Prove Th. 5.3.2.14.

PROOF:

Let $q(x)$ be the monic minimal polynomial of a over R . Then, by Prob. 20 above, there exists a polynomial $g(x)$ over R such that

$$f(x) = g(x)q(x) \quad (1)$$

where $f(x)$ is monic. Further, by Th. 5.2.1.11,

$$f(x) = c_f g^*(x) q^*(x) \quad (2)$$

where $g^*(x)$ and $q^*(x)$ are primitive, while $c_f \in R$, $c_f > 0$ and $q(x) = c_q q^*(x)$. Then $c_f = 1$, since $f(x)$ is monic, hence primitive.

Since $q^*(x)$ and $g^*(x)$ have integral coefficients, by Df. 5.2.1.10, and since their product $f(x)$ is monic, they must be monic. Hence $c_q = 1$, since $q(x)$ is also monic by hypothesis. Thus $q(x) = q^*(x)$, which implies that every coefficient of $q(x)$ is a rational integer, and the number a which satisfies $q(x)$ must be an algebraic integer, by Df. 5.3.2.13, completing the proof.

22. If a is an algebraic number, then there exists a rational integer k such that ka is an algebraic integer.

PROOF:

By Df. 5.3.2.2, a satisfies an equation

$$c_n x^n + c_{n-1} x^{n-1} + \cdots + c_0 = 0$$

where $c_i \in I$, $i=0,1,\dots,n$, which implies that $c_n a$ satisfies an equation

$$x^n + c_{n-1} x^{n-1} + c_n c_{n-2} x^{n-2} + c_n^2 c_{n-3} x^{n-3} + \cdots + c_n^{n-1} c_0 = 0$$

Let $k = c_n$, and the proof is complete.

23. If a satisfies an equation

$$f(x) = x^n + c_{n-1} x^{n-1} + c_{n-2} x^{n-2} + \cdots + c_0 = 0 \quad (1)$$

where the c_k , $k=0,1,\dots,n-1$, are algebraic integers, then a is also an algebraic integer.

PROOF:

If the $c_k^{(i)}$ are the conjugates of c_k over R , then, by Prob. 15 of §5.2.2, the product over the conjugates,

$$p(x) = \prod (x^n + c_{n-1}^{(i_1)} x^{n-1} + c_{n-2}^{(i_2)} x^{n-2} + \cdots + c_0^{(i_n)}) \quad (2)$$

has rational coefficients, which in this case must be also rational integral coefficients, since they are algebraic integers as they have been introduced by adding the products of the $c_k^{(i)}$.

Furthermore, $p(a)=0$, since $f(x) \mid p(x)$. Also, as is quite obvious in (2), $p(x)$ is monic. Hence, by Th. 5.3.2.14, a is an algebraic integer.

24. Prove Th. 5.3.2.15.

PROOF:

- (i) Since $p = a + b\sqrt{m}$ yields $N(p) = a^2 - b^2m$ and $T(p) = 2a$, i.e. $N(p), T(p) \in I$, it follows that $2a = c \in I$ implies

$$(2b)^2m = c^2 - 4N(p) \quad (1)$$

which is again a rational integer. Since m has no square factor, by hypothesis, $2b$ must be a rational integer. Hence let $2b = d$, and

$$p = (c + d\sqrt{m})/2 \quad (2)$$

where $c, d \in I$. Then, by (1),

$$d^2m \equiv c^2 \pmod{4} \quad (3)$$

If c is not a multiple of 4, then $c^2 \equiv 1 \pmod{4}$, which implies

$$d^2m \equiv 1 \pmod{4}$$

Hence d is not a multiple of 4 and $d^2 \equiv 1 \pmod{4}$, which implies

$$m \equiv 1 \pmod{4}$$

Hence $m \not\equiv 1 \pmod{4}$ implies that d must be a multiple of 2, which in turn implies $a, b \in I$.

- (ii) If $m \equiv 1 \pmod{4}$, then, from (3),

$$d^2 \equiv c^2 \pmod{4}$$

which implies $4 \mid (d-c)(d+c)$, which in turn implies $2 \mid (c-d)$ or $2 \mid (c+d)$.

If $2 \mid (c+d)$, then $2 \mid (c-d)$, since $c-d = (c+d) - 2d$, which is a multiple of 2. Hence

$$c \equiv d \pmod{2}$$

Conversely, if $c \equiv d \pmod{2}$, it must be the case that

$$T((c + d\sqrt{m})/2) = c \in I$$

and

$$N((c + d\sqrt{m})/2) = (c^2 - d^2m)/4$$

which is again a rational integer, since $c^2 \equiv d^2 \pmod{4}$ and $m \equiv 1 \pmod{4}$ imply $c^2 \equiv d^2m \pmod{4}$.

Hence $(c + d\sqrt{m})/2$ is an algebraic integer, which completes the proof.

Supplementary Problems

Part 5

- 5.1. Subtraction in the rational number field R is well-defined; so is division in R .
- 5.2. $A \subset B$ for every $A, B \subset C$ (cf. Df. 5.1.2.3) iff there exists some $C \subset C$ such that $A + C = B$.
- 5.3. If p is a positive rational number greater than 1, then there exists an element s in a D -cut S such that $p \cdot s \in S'$.
- 5.4. Generalize Prob. 36 of §5.1.2.

5.5. Any nonempty subset S of the ordered field \bar{R} of real numbers which has an upper bound in \bar{R} has a l.u.b. in \bar{R} .

5.6. The rational number field R is not complete.

*5.7. Any Archimedean-ordered field \bar{F} is isomorphic to a subfield \bar{R}' of the real number field \bar{R} , where the isomorphic mapping is unique. Furthermore, the order-isomorphism of \bar{F} into \bar{R} is unique; viz. it is an identity mapping.

5.8. If a field F is Archimedean-ordered and contains the rational number field R , then there exists an integer m such that, for some integer k ,

$$mn^k \leq a \leq (m+1)n^k$$

where $a \in F$, $n \in N$ and $n > 1$.

5.9. If F is an Archimedean-ordered field which contains R , and if $a \in F$ is positive, then there exists a natural number k such that

$$1/n^k < a$$

where $n > 1$ is a natural number.

5.10. If a_1, a_2, \dots, a_n are positive numbers, then

$$(a_1 + a_2 + \dots + a_n)/n \geq \sqrt[n]{a_1 a_2 \cdots a_n}$$

where equality holds iff $a_1 = a_2 = \dots = a_n$.

5.11. If a_1, a_2, \dots, a_n are positive and $s = a_1 + a_2 + \dots + a_n$, then

$$(s - a_1)(s - a_2) \cdots (s - a_n) \geq (n-1)^n a_1 a_2 \cdots a_n$$

5.12. If a_1, a_2, \dots, a_n are all distinct and positive, and if $s = a_1 + a_2 + \dots + a_n$, then

$$s/(s - a_1) + s/(s - a_2) + \dots + s/(s - a_n) > n^2/(n-1)$$

5.13. $4x^3 - 2x^2 + x + 1$ is irreducible over R .

5.14. If a is a prime integer, then $x^n - a$ is irreducible over R .

5.15. If $a \in R$ and there exists no p th root of a in R , where p is a positive prime, then $x^n - a$ is irreducible over R .

5.16. $x^3 + y^3 + z^3 - 3xyz$, where $c \in C$, is irreducible over C iff $c^3 = 1$.

*5.17. If p is a prime and n a positive integer greater than 1, then

$$x^{p^{n-1}(p-1)} + x^{p^{n-2}(p-2)} + \dots + x^{p^{n-1}} + 1$$

is irreducible over R .

*5.18. Let $f(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_0$, where $p \nmid a_n$, $p \mid a_{n-1}, \dots, p \mid a_0$, for a prime p . If $p^2 \nmid a_0$, then $f(x)$ is irreducible over R .

*5.19. Let a polynomial with integral coefficients be

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_0$$

where $p^2 \nmid a_0$, $p \nmid a_r$, $0 < r \leq n$, but $p \mid a_{r-1}, p \mid a_{r-2}, \dots, p \mid a_0$, for a prime p . If $f(x)$ can be decomposed, say,

$$f(x) = g(x)h(x)$$

then at least one of $g(x)$ and $h(x)$ is of degree greater than r .

*5.20. Prove, by Supplementary Problem 5.19 above, that $x^5 + x^4 + 10x^3 + 4x - 6$ is irreducible over R .

*5.21. If a polynomial $f(x)$ with integral coefficients can be decomposed over R , say,

$$\begin{aligned} f(x) &= a_r x^r + a_{r-1} x^{r-1} + \cdots + a_0 \\ &= (b_s x^s + b_{s-1} x^{s-1} + \cdots + b_0)(c_t x^t + c_{t-1} x^{t-1} + \cdots + c_0) \end{aligned}$$

where $r = s + t$ and $s \geq t \geq 1$, and $p \nmid a_n$, but $p \mid a_{n-1}, \dots, p \mid a_0$, then

$$p^2 \mid a_{t-1}, p^2 \mid a_{t-2}, \dots, p^2 \mid a_0$$

*5.22. Find, by Supplementary Problem 5.21 above, an integer k for which $f(x) = x^5 + 5x + 5k$ is irreducible over I .

5.23. Given the three roots a, b, c of $x^3 + px^2 + qx + r = 0$, find an equation whose roots are a^3, b^3, c^3 .

5.24. If the three roots of $x^3 + qx + r = 0$ are a, b, c , then the six roots of $r^2(x^2 + x + 1)^3 + q^3x^2(x + 1)^2 = 0$ are $b/a, c/a, a/b, c/b, a/c, b/c$.

5.25. Given an equation with real coefficients,

$$f(x) = x^n + p_1 x^{n-1} + p_2 x^{n-2} + \cdots + p_n = 0 \quad (1)$$

and a root r of (1), find (i) an equation whose root is $1/r$, (ii) an equation whose root is $-r$, (iii) an equation whose root is kr , (iv) an equation whose root is $r + m$.

5.26. If $a_0 x^n + a_1 x^{n-1} + \cdots + a_n = 0$ is a reciprocal equation, then either

$$\begin{aligned} a_0 &= a_n, & a_1 &= a_{n-1}, & a_2 &= a_{n-2}, \dots, \\ \text{or} & & a_0 &= -a_n, & a_1 &= -a_{n-1}, & a_2 &= -a_{n-2}, \dots \end{aligned}$$

5.27. If $x^2 + ax + b = 0$, where $a, b \in I$, has roots in R , then they are in I .

5.28. Solve $x^3 - 9x^2 + 36x - 48 = 0$.

5.29. Solve $x^4 - 17x^2 - 34x - 30 = 0$.

5.30. Determine the discriminant of (i) the cubic equation $x^3 + px + q = 0$,
(ii) the quartic equation $x^4 + px^2 + qx + r = 0$.

5.31. Find the necessary and sufficient condition that the polynomial

$$ax^3 + 3bx^2 + 3cx + d$$

be transformed to

$$p(x-u)^3 + q(x-v)^3 + r(x-w)^3 = 0$$

where u, v, w are the three roots of $cx^3 + 3fx^2 + 3gx + h = 0$.

5.32. If $a^{11} = 1$, then find a quintic equation whose roots are $a + a^{10}, a^2 + a^9, a^3 + a^8, a^4 + a^7, a^5 + a^6$.

*5.33. If $x_1 = a, x_2 = a + d, \dots, x_{n-1} = a + ((n-1) - 1)d$, and $x_n = a + (n-1)d$ are the n roots of

$$x^n + p_1 x^{n-1} + p_2 x^{n-2} + \cdots + p_0 = 0$$

then there exists the following relation among roots:

$$d = (2/n) \sqrt{3((n-1)p_1^2 - 2np_2)/(n^2 - 1)}$$

and

$$a = -(p_1/n) - ((n-1)d)/2$$

5.34. If $1, a_1, a_2, \dots, a_{n-1}$ are the n roots of $x^n - 1 = 0$, then find an equation whose roots are $1 - a_1, 1 - a_2, \dots, 1 - a_{n-1}$.

5.35. If an equation

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0 = 0, \quad a_0, a_1, \dots, a_n \in I$$

has an integral root r , then for some $m \in I$, $(r - m) \mid f(m)$.

5.36. Find, by Supplementary Problem 5.35 above, an integral root of

$$x^5 - 11x^4 + 65x^3 - 22x^2 - 143x - 507 = 0$$

5.37. Given a quartic equation $x^4 + ax^3 + bx^2 + cx + d = 0$, the product of two roots equals the product of two other roots iff $a^2d - c^3 = 0$.

5.38. If f is a polynomial in x_1, x_2, \dots, x_n , and if it is symmetric over R , so is f^3 .

5.39. A polynomial $f(x, y, z)$ can be expressed as

$$f_n(x, y)z^n + f_{n-1}(x, y)z^{n-1} + \dots + f_1(x, y)z + f_0(x, y)$$

where each of the $f_k(x, y)$, $k = 0, 1, \dots, n$, is a polynomial in x and y .

5.40. Any multiple algebraic extension of a field is a simple algebraic extension.

*5.41. If a polynomial $f(x)$ of degree n over a field F is irreducible over F , but reducible over an extension $F[\sqrt[n]{a}]$ of F , where $a \in F$ and $\sqrt[n]{a} \notin F$, then n is an even number and $f(x)$ can be decomposed into two factors of degree $n/2$ which are irreducible over $F[\sqrt[n]{a}]$.

5.42. An algebraic extension E , which is created by adjoining $\sqrt{-3}$ to $R[\sqrt[3]{2}]$, is of degree 6 and coincides with an extension $E' = R[\sqrt[3]{2} + \sqrt{-3}]$.

5.43. The set of all numbers of the form $a + b\sqrt{3}$, where $a, b \in R$, forms an algebraic number field, and any number field which contains $\sqrt{3}$ consists of the numbers of the same form.

5.44. If $p, q \in R[\sqrt{2}]$ and $q \neq 0$, then there exist $r, s \in R[\sqrt{2}]$ such that

$$p = sq + r, \quad \text{where } |N(r)| < |N(q)|$$

5.45. If $m \equiv 1 \pmod{4}$, then every algebraic integer in $R[\sqrt{m}]$ is of the form

$$a + b((1 + \sqrt{m})/2), \quad \text{where } a, b \in R$$

5.46. Every rational integer is an algebraic integer in $R[\sqrt{m}]$, and conversely, any integer in $R[\sqrt{m}]$ which is a rational number is a rational integer.

5.47. Every element r in $R[\sqrt{m}]$ is an algebraic integer if there exists a nonzero algebraic integer s such that

$$sr, sr^2, sr^3, \dots$$

are also algebraic integers.

5.48. If g is a Gaussian integer, then $N(g) = 0$ if $g = 0$, $N(g) = 1$ if $g = \pm 1$ or $\pm i$, and otherwise $N(g) > 1$.

5.49. If a and b are relative primes in the domain \bar{G}^* of all Gaussian integers, i.e. if a and b have no common divisors except units, then there exist $c, d \in \bar{G}^*$ such that $ac + bd = 1$.

5.50. If $a, b, c \in \bar{G}^*$, where \bar{G}^* is the integral domain of all Gaussian integers, and if $a \mid bc$ when a and b are relatively prime, then $a \mid c$.

5.51. The unique factorization theorem holds in \bar{G}^ .

5.52. The set \bar{I} of all algebraic integers forms an integral domain.

Answers and Hints to Supplementary Problems

Part 1

TAUTOLOGIES

- 1.1. Cf. §1.1.1, Prob. 6ff.
 1.2. Use truth-tables (cf. §1.1.1, Prob. 8).
 1.3. Cf. §1.1.1, Prob. 12.
 1.4. By truth-tables, or by deduction (cf. §1.1.1, Prob. 15-18).
 1.5. $p \rightarrow q \equiv p \mid (q \mid q), \quad p \rightarrow q \equiv ((p \downarrow p) \mid q) \downarrow ((p \mid p) \downarrow q);$
 $p \leftrightarrow q \equiv \{(p \mid (q \mid q)) \mid (q \mid (p \mid p))\} \mid \{(p \mid (q \mid q)) \mid (q \mid (p \mid p))\},$
 $p \leftrightarrow q \equiv [\{((p \downarrow p) \downarrow q) \downarrow ((p \downarrow p) \downarrow q)\} \downarrow \{((p \downarrow p) \downarrow q) \downarrow ((p \downarrow p) \downarrow q)\}]$
 $\quad \downarrow [\{((q \downarrow q) \downarrow p) \downarrow ((q \downarrow q) \downarrow p)\} \downarrow \{((q \downarrow q) \downarrow p) \downarrow ((q \downarrow q) \downarrow p)\}].$
 (Note. Check the validity of this answer by truth-tables; cf. §1.1.1, Prob. 2.)
 1.6. $p \mid q \equiv ((p \downarrow p) \downarrow (q \downarrow q)) \downarrow ((p \downarrow p) \downarrow (q \downarrow q)),$
 $p \downarrow q \equiv ((p \mid p) \mid (q \mid q)) \mid ((p \mid p) \mid (q \mid q)).$
 (Note. As above, check the validity of this answer by truth-tables.)
 1.7. Cf. §1.1.1, Prob. 6, (iii).
 1.8. $\{a \rightarrow (a(\overline{b \vee c}))\} \rightarrow (\overline{a} \vee \overline{bc}).$
 1.9. Cf. §1.1.1, Prob. 12 (break the negation lines as often as possible until the right-hand side of the tautology is obtained).
 1.10. Cf. §1.1.1, Prob. 11, then Prob. 15-18.
 1.11. As above.
 1.12. (i) False. (ii) True. (iii) False.
 1.13-15. Cf. §1.1.1, Prob. 15-18.

QUANTIFICATIONS

- 1.16-19. Cf. §1.1.2, MTh. 1.1.2.6-7, Prob. 2ff.
 1.20. Proceed the proof indirectly, starting with Hyp., i.e.,

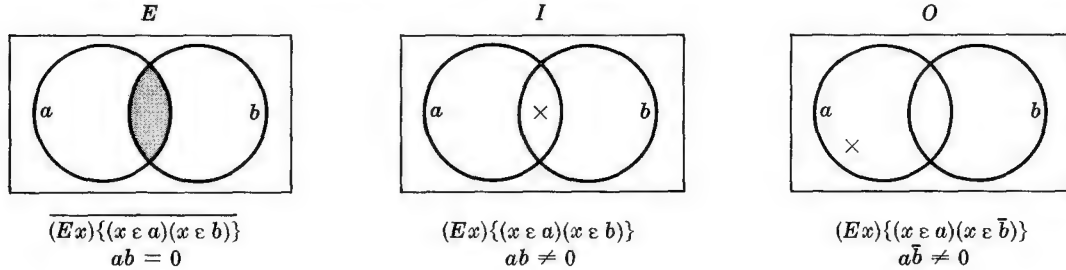
$$(x)\{I(x) \rightarrow (O(x) \rightarrow I(x^2) \overline{O(x^2)})\}$$
 where I and O predicate an integer and an odd integer respectively, and denying the conclusion, i.e.,

$$(x)\{I(x) \rightarrow (I(x^2) O(x^2) \rightarrow O(x))\}$$
 If the last step is of the form $p\overline{p}$, then the proof is complete.

Part 2

- 2.1. Cf. Df. 2.1.7 and, by indirect method, assume that, for every set S , " $\emptyset \subset S$ " is false, which implies the existence of a set S' such that $\emptyset \not\subset S'$, which in turn implies, by Df. 2.1.1b and Df. 2.1.2, an element x such that $x \notin S'$ and $x \in \emptyset$, contradicting Df. 2.1.7.
 2.2. Cf. §2.3, Prob. 7.
 2.3. (i) $A \cap (B \cap C) = (A \cap B) \cap C,$ (ii) $A \cap B = B \cap A,$ (iii) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$
 (Cf. §2.3, Prob. 7.)

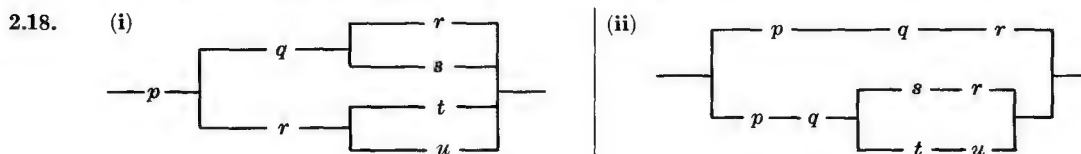
- 2.5. Let $x \in B$; then the hypothesis $A \cap B = \emptyset$ implies $x \notin A$. But also $x \in C$, since obviously $B \subset C$; i.e. $x \notin A$ and $x \in C$, meaning $x \in C - A$. Hence $B \subset (C - A)$. Likewise, $(C - A) \subset B$, and altogether $B = C - A$.
- 2.6. If $a = c$ and $c = d$, then trivially, $\{\{a\}, \{a, b\}\} = \{\{c\}, \{c, d\}\}$. Conversely, since $\{a\} = \{a\} \cap \{a, b\}$ and $\{a, b\} = \{a\} \cup \{a, b\}$, and since, by hypothesis, $\{a\} = \{a\} \cap \{a, b\} = \{c\} \cap \{c, d\} = \{c\}$, it follows that $a = c$. If, furthermore, $b \neq c$, then $b \in \{c, d\}$, and this implies $b = d$. If, finally, $b = c$, then $a = b = c = d$.
- 2.7. (i) $X \times X = \{(p \cdot p), (p, q), (q, p), (q, q)\}$, and likewise (ii).
(iii) $X \times Y = \{(p, r), (p, s), (p, t), (q, r), (q, s), (q, t)\}$, and likewise (iv).
- 2.8. E.g. for A : $(x)\{f(x) \rightarrow g(x)\} \equiv (x)\{(x \in a) \rightarrow (x \in b)\} \equiv \overline{(Ex)\{(x \in a)(x \in \bar{b})\}} \equiv \overline{(Ex)(x \in a\bar{b})}$. Hence $a\bar{b} = 0$. Justify these steps, then do likewise for $E.I.O.$
- 2.9.



- 2.10. (i) $p \vee (p \wedge (p \vee r)) \vee q \wedge (p \vee q') = p \vee (p \wedge (p \vee r)) \vee ((q \wedge p) \vee (q \wedge q'))$
 $= p \vee (p \wedge (p \vee r)) \vee ((q \wedge p) \vee 0) = p \vee (p \wedge (p \vee r)) \vee (q \wedge p)$
 $= p \vee p \vee (q \wedge p) = p \vee (q \wedge p) = p$
- (Specify which one of B1-6 has been employed to justify each step taken above, referring to Df. 2.4.2.1.) Do likewise for the rest;
- (ii) 0, (iii) $(p \wedge q) \vee (p' \wedge q' \wedge s' \wedge t)$.
- 2.11. Cf. §1.1.1, Prob. 7 note and Prob. 8; also Prob. 15-18.
- 2.12. Cf. Prob. 2.10 above; (i) $(p \wedge q) \vee (p' \wedge q') \vee r$, (ii) $p \vee q$, (iii) $p \wedge q$.
- 2.13. $o(A) = o(A \cap B) + o(A \cap B')$, since $(A \cap B) \cap (A \cap B') = \emptyset$ and $A = (A \cap B) \cup (A \cap B')$. Likewise, $o(B) = o(A \cap B) + o(A' \cap B)$. Adding these equations,
- $$o(A \cap B') + o(A' \cap B) = o(A) + o(B) - 2o(A \cap B)$$
- But, since $(A \cap B) \cap (A \cap B') \cap (A' \cap B) = \emptyset$ and also
- $$A \cup B = (A \cap (B \cup B')) \cup (B \cap (A \cup A')) = (A \cap B) \cup (A \cap B') \cup (A \cap B) \cup (A' \cap B)$$
- $$= (A \cap B) \cup (A \cap B') \cup (A' \cap B)$$
- which implies $o(A \cup B) = o(A \cap B) + o(A \cap B') + o(A' \cap B)$, a substitution yields the conclusion. (The reader should notice, then prove, a more generalized form:
- $$o(A \cup B \cup C) = o(A) + o(B) + o(C) - o(A \cap B) - o(B \cap C) - o(C \cap A) + o(A \cap B \cap C)$$
- Further generalization can be carried out, of course.)
- 2.14. By hypothesis, $B = B \cap (A_1 \cup A_2 \cup \dots \cup A_n) = (B \cap A_1) \cup (B \cap A_2) \cup \dots \cup (B \cap A_n)$, which immediately yields the theorem.
- 2.15. Let R_1 and R_8 be given as in the text, and
- $$\begin{aligned}
 R_2 &= (x \in A) \wedge (x \in B) \wedge (x \notin C) & R_5 &= (x \notin A) \wedge (x \in B) \wedge (x \in C) \\
 R_3 &= (x \in A) \wedge (x \notin B) \wedge (x \in C) & R_6 &= (x \notin A) \wedge (x \in B) \wedge (x \notin C) \\
 R_4 &= (x \in A) \wedge (x \notin B) \wedge (x \notin C) & R_7 &= (x \notin A) \wedge (x \notin B) \wedge (x \in C)
 \end{aligned}$$
- then
- $$\begin{aligned}
 \text{(i)} \quad U &= R_1 \cup R_2 \cup \dots \cup R_8 & \text{(vi)} \quad A \cup B &= R_1 \cup R_2 \cup R_3 \cup R_4 \cup R_5 \cup R_6 \\
 \text{(ii)} \quad A &= R_1 \cup R_2 \cup R_3 \cup R_4 & \text{(vii)} \quad A \cap B &= R_1 \cup R_2 \\
 \text{(iii)} \quad B &= R_1 \cup R_2 \cup R_5 \cup R_6 & \text{(viii)} \quad A' \cap (A \cap B) &= \emptyset \\
 \text{(iv)} \quad C &= R_1 \cup R_3 \cup R_5 \cup R_7 & \text{(ix)} \quad (A \cup B) \cap C &= R_1 \cup R_3 \cup R_5 \\
 \text{(v)} \quad A' &= R_5 \cup R_6 \cup R_7 \cup R_8 & \text{(x)} \quad (A \cap B') \cap C' &= R_4
 \end{aligned}$$
- 2.16. (i) 10, (ii) 20

- 2.17. $a + b + c - d - e - f + g$, where $g = o(A \cap B \cap C)$ (since
- $$\begin{aligned} o(A \cup B \cup C) &= o(A \cup B) + o(C) - o((A \cup B) \cap C) \\ &= o(A) + o(B) - o(A \cap B) + o(C) - o((A \cap C) \cup (B \cap C)) \\ &= o(A) + o(B) + o(C) - o(A \cap B) - (o(A \cap C) + o(B \cap C) - o(A \cap B \cap C)) \\ &= o(A) + o(B) + o(C) - o(A \cap B) - o(A \cap C) - o(B \cap C) + o(A \cap B \cap C) \end{aligned}$$

cf. Prob. 2.13 above and also Prob. 2.19, (iii), below).



- 2.19. Verbally, (i) if an event x is impossible to occur, it has probability 0; note that the converse does not hold, since it cannot be said that x is impossible to occur if $P(x) = 0$. (ii) If x is any event, it has any probability between absolute impossibility and absolute certainty. (iii) The probability that at least one of two events x and y occurs is the sum of the probability that x occurs and the probability that y occurs, from which the probability that both x and y occur is subtracted.
- 2.20. As above, verbally, (iv) if x and y are mutually exclusive events, then the probability that x or y occurs, i.e. at least one of x and y occurs, is the sum of their individual probabilities. (v) Either an event occurs or it does not; i.e. the probability that an event x does not occur is the difference between 1 (certainty) and the probability that x does occur.

Part 3

- 3.1. There are four (exhaustive and mutually exclusive) types A_1, A_2, A_3, A_4 of symmetries with respect to the four different axes of symmetries, viz.,
- (i) A_1 : two opposite vertices;
 - (ii) A_2 : two centers of opposite faces;
 - (iii) A_3 : one vertex and the center of its opposite edge;
 - (iv) A_4 : two centers of opposite edges.
- The regular tetrahedrons, hexahedrons, octahedrons, dodecahedrons, and icosahedrons have, respectively, $A_1 = 0, 4, 3, 10, 6$, $A_2 = 0, 3, 4, 6, 10$, $A_3 = 4, 0, 0, 0, 0$, $A_4 = 3, 6, 6, 15, 15$.
- 3.2. There are three exhaustive and mutually exclusive types of rotations:
- (i) R_1 , with respect to A_1 (cf. Prob. 3.1 above), which yields, counting in the same order as above, 3, 3, 4, 3, 5, respectively;
 - (ii) R_2 , dually with respect to A_2 and A_3 , which yields, respectively, 3, 4, 3, 5, 3;
 - (iii) R_3 , with respect to A_4 , where $R_3 = A_4$. Hence, counting the original position as 1, the total number is

$$R = 1 + A_1(R_1 - 1) + A_2(R_2 - 1) + A_3(R_2 - 1) + A_4$$

and R of the regular octahedron, for instance, is

$$1 + 3(4 - 1) + 4(3 - 1) + 0(3 - 1) + 6 = 24$$

- 3.3. By C2-3, $ab = a(be) = (ba)e = ba$, yielding G5, which in turn implies $a(bc) = (ba)c = (ab)c$, proving G2; the rest follows immediately, completing the proof.
- 3.4. Construct the multiplication table of the given functions under the prescribed operative rule.
- 3.5. (i) By cancellation law. (ii) By induction, $(bab^{-1})^{k+1} = (bab^{-1})^k bab^{-1} = ba^k b^{-1} bab^{-1} = ba^{k+1} b^{-1}$. (If $n < 0$, then $bab^{-1} ba^{-1} b^{-1} = e$, which implies $(bab^{-1})^{-1} = ba^{-1} b^{-1}$, hence $(bab^{-1})^n = (ba^{-1} b^{-1})^{-n} = b(a^{-1})^{-n} b^{-1} = ba^n b^{-1}$.) (iii) By induction, as in (ii).
- 3.6. Cf. Th. 3.1.2.6, (ii).

- 3.7. Actual transpositions: $(a\ b)$, $(b\ c)$, $(c\ a)$ in (i), (ii), (iii) leave them invariant; hence, by definition, the proof is complete.
- 3.8. Cf. Prob. 3.7 above.
- 3.9. Since the product of a symmetric polynomial and an alternating polynomial yields the change of signs by a transposition, it is by definition an alternating polynomial; likewise, by definition, the product of two alternating polynomials is a symmetric polynomial, since the sign is unchanged here by transpositions.
- 3.10. In general $(1\ 2\ a)(1\ 2\ a)(1\ 2\ b) = (1\ a\ b) = (1\ a)(1\ b)$, which evidently belongs to G if $a \neq 2$ and $b \neq 2$. Also $a \neq 2$ implies $(1\ 2)(1\ a) = (1\ 2\ a)$ and $(1\ a)(1\ 2) = (1\ a\ 2) = (1\ a\ 2)(1\ 2\ a)$; hence $(1\ 2)(1\ a)$ and $(1\ a)(1\ 2)$ belong to G and every even permutation belongs to G . If G contains even one odd permutation, then every odd permutation is the product of itself and an even permutation. Hence G is an alternating group or a symmetric group (cf. Df. 3.1.2.16, 18).
- 3.11. Cf. Prob. 3.10 above.
- 3.12. Cf. Th. 3.2.1.4.
- 3.13. Cf. Prob. 3.12 above.
- 3.14. Cf. Prob. 3.13 above and Th. 3.2.1.2.
- 3.15. Since, by hypothesis, there exist two integers p and q such that $d = pm + qn$, it follows
- $$(g^d)^s = (g^{pm+qn})^s = (g^m)^{ps} (g^n)^{qs}$$
- and also since there exist m' and n' such that $m = dm'$ and $n = dn'$, it follows
- $$(g^m)^u (g^n)^v = g^{dm'u + dn'v} = (g^d)^{m'u + n'v}$$
- 3.16. By Th. 3.2.2.10, the orders of the proper subgroups of S_3 is either 2 or 3, which implies transpositions of either $(a\ b)$, $(b\ c)$, $(c\ a)$ or $(a\ b\ c)$, $(a\ c\ b)$, which in turn yield the following four subgroups:
- $$G_1: (1), (a\ b); \quad G_2: (1), (b\ c); \quad G_3: (1), (c\ a); \quad G_4: (1), (a\ b\ c), (a\ c\ b)$$
- 3.17. The transpositions which leave the polynomial as it is are: (1) , $(a\ b)$, $(c\ d)$, $(a\ b)(c\ d)$, and the transpositions which interchange the terms of the polynomial are: $(a\ c)(b\ d)$, $(a\ d)(b\ c)$, $(a\ d\ b\ c)$, $(a\ c\ b\ d)$, which altogether do form a subgroup of S_4 . In either case the polynomial remains unchanged.
- 3.18. Cf. Prob. 3.17 above.
- 3.19. This is a direct result from Prob. 3.18 above, since the number of permutations belonging to S_4 is $4! = 24$, which is exhausted, mutually exclusively, by M , $M(bc)$, and $M(bd)$.
- 3.20. E.g. $(ac) = (bc)^{-1}(ab)(bc)$, etc.
- 3.21(a). As can be readily verified, the subgroup O_1 of the octahedral group O corresponds to the set of 4 rotations with A_1 fixed (cf. the figure at right). Also, in general, if P_i represents a rotation which moves A_1 to A_i , $i = 1, 2, \dots, 6$, then every rotation which belongs to the right-coset O_1P_i moves A_1 to A_i , and conversely (since QP_i^{-1} , where Q is the rotation which moves A_1 to A_i , moves A_1 to A_1 and consequently belongs to O_1 , which implies $Q \in O_1P_i$). Hence

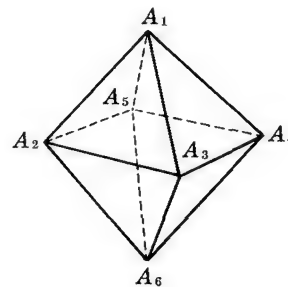
$$O = O_1P_1 \cup O_1P_2 \cup \dots \cup O_1P_6$$

which implies the order of O is 4 (i.e. the order of O_1) times 6 (i.e. the number of right-cosets), i.e. 24.

- 3.21(b). Let V_i , $i = 1, 2, \dots, 5$, be the number of the vertices of the regular tetrahedron, hexahedron, octahedron, dodecahedron, and icosahedron, respectively, i.e. 4, 6, 8, 12, 20, and P be the number of edges which go through a vertex in rotations (cf. Prob. 3.21 above); then the set of P_i , $i = 1, 2, \dots, 5$, represents the orders of the subgroups formed by rotations with respect to the given regular polygons, respectively, which also corresponds to the set R_i of Prob. 3.2; i.e. 3, 3, 4, 3, 5. The total number R of rotations, in this context, is then given by

$$R = P_i V_i, \quad i = 1, 2, \dots, 5$$

yielding 12, 24, 24, 60, 60 for the five regular polygons, respectively.



- 3.22. Cf. §3.1.2, Prob. 5 and Fig. 3.1.2c, where 0 and 7 forms a right coset C , while 1 and $7 \cdot 1$, 2 and $7 \cdot 2$, 3 and $7 \cdot 3$, constitute the right-cosets $C1, C2, C3$, respectively. Since the order of D_4 is 8,

$$D_4 = C \cup C1 \cup C2 \cup C3$$

On the other hand, there are five conjugate classes with respect to D_4 ; i.e., in the same context, (i) 0, (ii) 1 and 3, (iii) 2, (iv) 4 and 5, (v) 6 and 7.

- 3.23. For every $a, b \in \bar{R}$, $ab \rightarrow |ab| = |a||b|$.

- 3.24. Use the n diagonals of the regular $2n$ -gon which connect the n pairs of opposite vertices.

- 3.25. If a and b are the generators of the cyclic groups of order m and n , then correspondence $a^i \rightarrow b^j$ is unique, since $a^i = a^j$ implies $i \equiv j \pmod{m}$, hence $i \equiv j \pmod{n}$, which in turn implies $b^i = b^j$. Thus $a^i a^j = a^{i+j} \rightarrow b^{i+j} = b^i b^j$ is the prescribed homomorphism.

- 3.26.

$$a \rightarrow \begin{pmatrix} g_1 G_1 & g_2 G_1 & \cdots & g_n G_1 \\ ag_1 G_1 & ag_2 G_1 & \cdots & ag_n G_1 \end{pmatrix}$$

$$\text{and } b \rightarrow \begin{pmatrix} g_1 G_1 & g_2 G_1 & \cdots & g_n G_1 \\ bg_1 G_1 & bg_2 G_1 & \cdots & bg_n G_1 \end{pmatrix} = \begin{pmatrix} ag_1 G_1 & ag_2 G_1 & \cdots & ag_n G_1 \\ abg_1 G_1 & abg_2 G_1 & \cdots & abg_n G_1 \end{pmatrix}$$

do imply a homomorphism:

$$ab \rightarrow \begin{pmatrix} g_1 G_1 & \cdots & g_n G_1 \\ abg_1 G_1 & \cdots & abg_n G_1 \end{pmatrix} = \begin{pmatrix} g_1 G_1 & \cdots & g_n G_1 \\ ag_1 G_1 & \cdots & ag_n G_1 \end{pmatrix} \begin{pmatrix} ag_1 G_1 & \cdots & ag_n G_1 \\ abg_1 G_1 & \cdots & abg_n G_1 \end{pmatrix}$$

- 3.27. Cf. Fig. 3.1.2e (§3.1.2, Prob. 8), where the tetrahedron has two types of rotations, viz.,

- (i) R_1 , with respect to the symmetric axis which goes through the vertices A, B, C, D : (1), (BCD) , (BDC) ; (1), (ACD) , (ADC) ; (1), (ABD) , (ADB) ; (1), (ABC) , (ACB) ;
 (ii) R_2 , with respect to the axes connecting the midpoints of AB and CD , AD and BC : (1), $(AB)(CD)$; (1), $(AD)(BC)$. Since these are exactly what A_4 represents, the isomorphism is established.

- 3.28. G_2 is the cyclic subgroup generated by c^m .

- 3.29. Since $g^{-1}cg \in g^{-1}G_1g = G_1$ for every element c which belongs to the normal subgroup G_1 of G , every element which belongs to the conjugate classes of c belongs also to G_1 .

- 3.30. Cf. §3.1.2, Prob. 4, 6, where (1) forms a subgroup; so does each of $\{(1), (12)\}$, $\{(1), (13)\}$, $\{(1), (23)\}$, $\{(1), (123), (132)\}$, and the last subgroup alone is the normal subgroup of S_3 . Let $A = \{(1), (13)\}$, for instance; then, since $(123)A = \{(123), (23)\}$ and $(132)A = \{(132), (12)\}$,

$$S_3 = \{A, (123)A, (132)A\} = \{A, A(123), A(132)\}$$

Also, since $(13)(1) = (13)$, $(13)(123) = (12)$, $(13)(132) = (23)$, it follows that $B = \{(1), (123), (132)\}$ implies $S_4 = \{B, (13)B\} = \{B, B(13)\}$; likewise

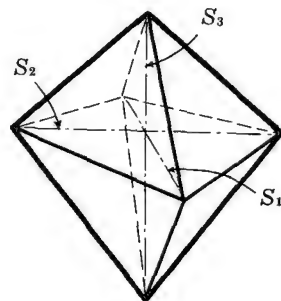
$$S_3 = \{B, (12)B\} = \{B, B(12)\}, \quad \text{and} \quad S_3 = \{B, (23)B\} = \{B, B(23)\}$$

- 3.31. Cf. §3.2.4, Prob. 9, (ii) (and also Prob. 3.21(a) above, giving reasons for "why not").

- 3.32. Let A_1, A_2, A_3 be moved to $A_1(R), A_2(R), A_3(R)$ by a rotation R which belongs to the octahedral group O ; then the correspondence

$$R \rightarrow \begin{pmatrix} A_1 & A_2 & A_3 \\ A_1(R) & A_2(R) & A_3(R) \end{pmatrix}$$

uniquely maps O onto S_3 of A_1, A_2, A_3 ; and another similar correspondence R' , together with RR' , establishes the desired homomorphism (cf. Prob. 3.26 above) of O onto the subset K of S_3 , corresponding to V_4 . Since the orders of O and S_3 are 24 and 6, respectively, the order of the kernel K is 4, by Th. 3.2.3.7 and Th. 3.2.6, 15, coinciding with that of V_4 , of course.



- 3.33. G/N is a cyclic group since it is of order p . Let one of its generators be Na ; then all cosets of N are N, Na, \dots, Na^{p-1} . If $c \in N$ and $c \neq e$, then $ca^i = a^i c$, since $a^{-i} Na^i = N$. Since every element of G can be expressed as either a^i or ca^i , while $a^i ca^i = ca^j a^i = ca^i a^j$, $ca^i ca^i = cca^i a^i = cca^i a^i = ca^i ca^j$, the proof is complete.

- 3.34. Since G has a subgroup G_1 of order 2, let the right-cosets be $G_1 a_1, G_1 a_2, G_1 a_3$, and let the permutation

$$\begin{pmatrix} G_1 a_1 & G_1 a_2 & G_1 a_3 \\ G_1 a_1 c & G_1 a_2 c & G_1 a_3 c \end{pmatrix}$$

correspond to $c \in G$, yielding a homomorphism H of G onto S_3 . If K is the kernel of G and $c \in K$, while $c \neq e$, then $G_1 a_1 = G_1 a_1 c$, which implies $a_1^{-1} G_1 a_1 = a_1^{-1} G_1 a_1 c$, which in turn implies $c \in a_1^{-1} G_1 a_1$. Likewise $c \in a_2^{-1} G_1 a_2$, $c \in a_3^{-1} G_1 a_3$. Hence $c \in G_1$, and $G_1 = \{e, c\}$, which implies $a_1^{-1} G_1 a_1 = a_2^{-1} G_1 a_2 = a_3^{-1} G_1 a_3$, which in turn makes G_1 a normal subgroup of order 2 and also commutative (cf. Prob. 3.33 above), contradicting the hypothesis. Hence $K = \{e\}$, and since G is of order 6, it is now isomorphic to S_3 .

- *3.35. Cf. Th. 3.2.7.9. *3.37. Cf. Prob. 3.36 above.
 *3.36. Cf. Df. 3.2.7.8a. *3.38. Cf. Th. 3.2.7.3.

Part 4

- *4.1. Consider two sets N_1 and N_2 which satisfy N1-4, establishing an isomorphism between them. (Note. Rings in general, on the other hand, offers many examples of an incomplete axiomatic system, for there do exist non-isomorphic rings, e.g. finite and infinite rings.)

- *4.2. Find, for each of N1-4, a set (or model) which does not satisfy it while it satisfies all the rest. (E.g. if N consists of three elements a, b, c such that

$$a' = b, \quad b' = c, \quad c' = a$$

then N1 cannot be satisfied here while N2-4 (or some equivalents) can.

- 4.3. Prove, first, that there exists at most one mapping which establishes a correspondence between every $a \in N$ and a number x_a , given $b \in N$, such that $x_1 = b'$ and also $x_a = (x_a)'$. Prove, then, the existence of a correspondence between every $a \in N$, given $b \in N$, and $a + b$ such that $b + 1 = b'$ and $b + a' = (b + a)'$ for every a and some b .
- 4.4-5. Cf. Prob. 4.3 above.
- 4.6. Cf. §4.1.2.3, Prob. 13.
- 4.7. If $N \subset S$, where every element of S is a difference between two elements of N , then S is a minimal ring, i.e. I , since any subring which contains N contains all differences in N (cf. Th. 4.1.1.7) and thus coincides with S . Conversely, if S is a minimal ring (i.e. I), then Prob. 18 of §4.1.1 leads to the desired conclusion.
- 4.8. Cf. Prob. 4.7 above, then consider two rings R_1 and R_2 which contain N , establishing an isomorphism between them.
- 4.9. The meet of all subrings of R which contain N is also a subring (cf. §4.2.1, Prob. 5), in fact a minimal subring; for it is contained in any subring which contains N , and this is indeed I .
- 4.10(a). Verify D1-11 with respect to I , in particular D11, and finally prove that $a \cdot 1 = a$ for any $a \in I$.
- 4.10(b). Let $a, b \in I$ and $a > b$; then $a - b = c$ is a positive integer (cf. Df. 4.1.2.2.5), i.e. $c \in N$. On the other hand, if $a, b \in N$, then $a = b + c$, which means $a > b$ in N (cf. Df. 4.1.2.3.2).
- 4.11. (i) $a < b + 1$ implies $x \in N$, where $x \geq 1$, such that $a + x = b + 1$, and if $x = 1$, then $a = b$, while if $x > 1$, then there exists $y \in N$ such that $x = y + 1$, which implies $(a + y) + 1 = a + (y + 1) = a + x = b + 1$, which in turn implies $a + y = b$, i.e. $a < b$.
 (ii)-(iii) By induction.
- 4.12. Cf. Df. 4.1.2.3.5; also Prob. 18-22 of §4.1.2.3.
- 4.13. Since $(a, b) = d$ divides $ax + by$ for any $x, y \in I$, $d \mid n$ if $ax + by = n$ has a solution at all (cf. §4.1.2.3, Prob. 32). Conversely, if $d \mid n$, then let $n = m'd$, yielding $n = n'(ax' + by') = a(n'x') + b(n'y')$, which reveals that $x = n'x'$ and $y = n'y'$ constitute a solution of the equation.

4.14. (i) Cf. §4.1.2.3, Prob. 33.

(ii) Prove, first, that $p \mid {}_p C_r$ (since ${}_p C_r = (p(p-1)\cdots(p-r+1))/r!$
 $= (p/r)((p-1)\cdots(p-r+1))/(r-1)!$ etc.).

Then, expanding by the Binomial Theorem,

$$(a+b)^p - a^p - b^p = {}_p C_1 a^{p-1} b + {}_p C_2 a^{p-2} b^2 + \cdots + {}_p C_{p-1} a b^{p-1}$$

where, since it is now known that $p \mid {}_p C_1$, etc., $p \mid ((a+b)^p - a^p - b^p)$. Likewise, letting $(a-b)^p = p^p n^p$ by hypothesis, $p^2 \mid (a^p - b^p)$.

4.15. Let $(a+b, a-b) = d$; then, since $(a, b) = 1$ by hypothesis and $d(a, b) = d = (da, db)$, it follows that $(a+b, a-b) = (da, db)$, implying $a+b = ad$ and $a-b = bd$, which yield $a = d(a+b)/2$, $b = d(a-b)/2$. Hence $2 \mid d$, and since $a+b$ and $a-b$ cannot be both simultaneously even, $d = 2$ (since if $d = 2k$, then $a = k(a+b)$, $b = k(a-b)$, and $(a, b) = k \neq 1$, contradicting the given hypothesis).

4.16. By induction and Df. 4.1.2.3.18.

4.17. By hypothesis $p \mid (a+b)(a-b) = a^2 - b^2$, which implies $p \mid (a-b)$ or $p \mid (a+b)$, i.e. $a \equiv b \pmod{p}$ or $a \equiv -b \pmod{p}$.

4.18. Several proofs are available (e.g. cf. Th. 3.2.6.13) for this Fermat's theorem, and the following is one of the most elementary and direct approach: by the Multinomial Theorem,

$$\underbrace{(x+y+\cdots)^p}_a = \underbrace{x^p + y^p + \cdots}_a + \sum (p!/(m!n!\cdots))x^m y^n \cdots$$

where $p \mid \sum (p!/(m!n!\cdots))$, and letting $x = y = \cdots = 1$,

$$a^p = \underbrace{1+1+\cdots+1}_a + \sum (p!/(m!n!\cdots)) = a + kp, \quad \text{i.e. } a^p - a = kp$$

where $p \mid (a^{p-1} - 1)$ since $(a, p) = 1$.

4.19. By induction and Df. 4.1.2.3.18.

4.20. $x \equiv 5 \pmod{11}$.

4.21. $x \equiv 75 \pmod{88}$.

4.22. $x \equiv 67 \pmod{90}$.

4.23. (i) By the Binomial Theorem, $(a+b)^p = a^p + {}_p C_1 a^{p-1} b + \cdots + {}_p C_{p-1} a b^{p-1} + b^p$, where

$${}_p C_k = (p(p-1)\cdots(p-k+1))/k!, \quad k = 2, 3, \dots, p-1$$

which must be a multiple of p as a whole while its denominator is not divisible by p . Hence, by hypothesis,

$${}_p C_k a^{p-k} b^k = {}_p C_k e a^{p-k} b^k = 0, \quad \text{i.e. } (a+b)^p = a^p + b^p$$

(ii) Substitute $(a-b)$ for a in (i), and the desired result is immediately obtained.

(iii) Generalize (i) by induction.

4.24. Since a field F contains at least one subfield (e.g. itself), the meet M of all subfields of F is always obtainable, which is a subfield of F itself (cf. §4.2.1, Prob. 5). Let M' be a subfield of M and different from M ; then $M' \subset F$, yet $M \not\subset M'$, which is evidently a contradiction; M is thus a prime field.

Furthermore, if M'' is also a prime field of F , then $M \cap M'' = L \subset F$, while $L \subseteq M$ and $L \subseteq M''$, which imply that L is a subfield of both M and M'' , i.e. two prime fields. Hence $M = M'' = L$, proving its uniqueness.

4.25. Let $a \in F$ and $a \neq 0$; then, by Th. 4.1.2.4.13, there exists $n \in N$ such that $na > 0$ if $a > 0$, and $na \neq 0$ for any $n \in I$, $n \neq 0$, since $(-n)a = -na$. If $a < 0$, then $-a > 0$, and likewise $n(-a) = -na \neq 0$ for any $n \in I$, $n \neq 0$. In either case $a \neq 0$ and $n \neq 0$ imply $na \neq 0$.

*4.26-31. Cf. Th. 4.1.2.5.19.

4.32. By matrix multiplication,

$$\bar{i}^2 = \begin{bmatrix} 0-1 & 0+0 \\ 0+0 & -1+0 \end{bmatrix} = -I = \bar{j}^2 = \bar{k}^2$$

and
$$\bar{i}\bar{j} = \begin{bmatrix} 0-i & 0+0 \\ 0+0 & i+0 \end{bmatrix} = \begin{bmatrix} -i & 0 \\ 0 & i \end{bmatrix} = \bar{k}$$

while
$$\bar{j}\bar{i} = \begin{bmatrix} 0+i & 0+0 \\ 0+0 & -i+0 \end{bmatrix} = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix} = -\bar{k}$$

likewise, $\bar{j}\bar{k} = \bar{i}$, $\bar{k}\bar{j} = -\bar{i}$, $\bar{k}\bar{i} = \bar{j}$, $\bar{i}\bar{k} = -\bar{j}$, verifying Df. 4.1.3.1.3.

4.33. Cf. Df. 4.1.3.1.5, Df. 4.1.3.1.11, and carry out actual computations.

4.34. (i) $(-5/2, 1, -3/2)$, (ii) $(0, -13/4, 1/4)$, (iii) $2b$.

4.35. Assume $a(1, 2, 7) + b(-2, 5, 4) + c(-1, 4, 5) = 0$, then solve it for a, b, c , which in this case turn out to be not all zero; hence the vectors are linearly dependent, by Df. 4.1.3.2.4.

4.36. By Df. 4.1.3.2.4, 7.

4.37. By actual scalar and matrix multiplication in $f(A)$.

4.38. Let $x = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, and find the relation among a, b, c, d ; i.e. $x = \begin{bmatrix} a & b \\ -2a & -2b \end{bmatrix}$.

4.39. Any matrix of the form: $X = \begin{bmatrix} a' & 0 & 0 \\ 0 & b' & 0 \\ 0 & 0 & c' \end{bmatrix}$

4.40. Cf. Df. 4.1.3.2.19, 20, 21.

4.41. Cf. §4.1.3.2, Prob. 33.

4.42. (i)-(iv) Cf. §4.1.3.2, Prob. 38-41.

4.43. $|A| = -2$, $A^* = \begin{bmatrix} -9 & 2 & 1 \\ 2 & 2 & -2 \\ 1 & -2 & 1 \end{bmatrix}$, $|A^*| = 4$, $A^{-1} = -A^*/2$.

4.44. $\begin{bmatrix} 1/3 & -14/15 & 2/15 \\ -2/3 & -1/3 & -2/3 \\ 2/3 & 2/15 & -11/15 \end{bmatrix}$.

4.45. $AX = B \rightarrow A^{-1}(AX) = A^{-1}B \rightarrow X = A^{-1}B$, while $A(A^{-1}B) = (AA^{-1})B = B$. Hence $A^{-1}B$ uniquely satisfies $AX = B$. Likewise BA^{-1} uniquely satisfies $YA = B$. Hence $A^{-1}B = BA^{-1}$, i.e. $X = Y$, if $AB = BA$.

4.46. Let M be a minimal subfield of a field F ; then the elements of the form: ne , $n = 0, 1, 2, \dots$, belong to M . Furthermore, if $n \neq 0$, $x \in M$ such that $nex = me$ can be expressed as (n, m) and, likewise, $x \in M$ such that $nex = -me$ may be represented as $(n, -m)$. Consider, then, R as the quotient field whose elements are of the form: (p, q) , $p, q \in I$.

4.47. Cf. the following addition and multiplication table for I/E :

+	E	$1+E$
E	E	$1+E$
$1+E$	$1+E$	E

\times	E	$1+E$
E	E	E
$1+E$	E	$1+E$

4.48. Since $2 \cdot 3 = 6$, $2 \cdot 5 = 10$, and $6, 10 \in (2)$, it follows that 3 and 5 are also ring elements in I , which implies $(6, 10) \subseteq (2)$, while also $(2) \subseteq (6, 10)$ since 2 is of the form: $6x + 10y$ ($x = 2$ and $y = -1$, in this case).

4.49. Utilize the fact that every cyclic subgroup of a cyclic group is again cyclic.

4.50. Cf. Df. 4.2.3.7.

4.51-53. Cf. Df. 4.2.2.11.

*4.54-58. Cf. §4.2.3.

Part 5

- 5.1. Cf. Th. 4.1.2.4.6, and prove that there uniquely exists (x/y) , $x, y \in I$, such that (i) $(a/b) - (x/y) = (c/d)$, (ii) $(a'/b')/(x/y) = (c'/d')$, where $x/y \neq 0$.
- 5.2. If $A + C = B$, then it is trivially true that $A \subseteq B$ (since $a + c = b$, for $a \in A, b \in B, c \in C$, implies $a < b$ and $A \subseteq B$, while there exists some $a \in A$ for every $r \in C$ such that $a + r \in A'$, implying $A \neq B$). Conversely, if $A \subseteq B$, define C to be the set of all elements of the form: $c_2 - c_1$, where $c_1 < c_2$ and $c_1, c_2 \in A' \cap B$. Then, for every $a \in A$, $a + (c_2 - c_1) = c_2 - (c_1 - a) < c_2$, which implies $a + (c_2 - c_1) \in B$, i.e. $A + C \subseteq B$. Likewise, $B \subseteq A + C$ can be proved, and $A + C = B$.
- 5.3. Choose $n \in N$ such that $n > 1/s$, i.e. $ns > 1$ for some $s \in S$, then use Th. 5.1.2.2 by letting $r = (p-1)/n$.
- 5.4. By induction.
- 5.5. Cf. Df. 5.1.2.1 and Th. 5.1.2.13. (Note. \bar{R} is the only ordered field which satisfies this theorem.)
- 5.6. Cf. Df. 5.1.2.15, and verify the existence of sequences without limits over R (cf. also Prob. 24-26 of §5.1.2). Or, more directly, consider the set R' of all rational numbers less than, say e ($\approx 2.71\dots$), which then has an upper bound (e.g. 2); further, assume R' to have a l.u.b., say r . Then, by Th. 5.1.1.5, r is not even an upper bound of R' if $r < e$, and also if $r < e$, it is not the l.u.b., either.
- 5.7. Cf. Prob. 43-44 of §5.1.2.
- 5.8. By hypothesis $n^k > 0$, and also by hypothesis, since F is Archimedean ordered, there exist $r, s \in N$ such that $rn^k > a$, $sn^k > -a$, the latter of which implies $(-s)n^k < a$, which further implies that a set S of integers t such that $tn^k \leq a$ contains $-s$ and is consequently not empty. Also, since $t < r$, S is bounded above and contains the greatest integer m , excluding $m+1$ from S ; hence the conclusion.
- 5.9. Prove, by induction, $n^k > k$, where $n > 1$ by hypothesis, which implies $n^ka > 1$, since $a > 0$ and there exists $k \in N$ such that $ka > 1$.
- 5.10. By induction.
- 5.11. Use the result of Prob. 5.10 above.
- 5.12. Verify, first $n/(1/a_1 + \dots + 1/a_n) \leq \sqrt[n]{a_1 \dots a_n} \leq (a_1 + \dots + a_n)/n$ then replace a_1, \dots, a_n by $1/(s-a_1), \dots, 1/(s-a_n)$.
- 5.13. If it is not irreducible, then it must have rational roots which in this case must be $\pm 1, \pm 1/2, \pm 1/4$ (cf. Th. 5.2.3.8), none of which is a zero of the given polynomial, however.
- 5.14. If there exist a factor with the roots r_1, \dots, r_m , then $r = r_1 \dots r_m$ is rational and $r^n = a^m$, which implies that r must be a rational root of $x^n - a^m$, which is impossible if $m < n$, however.
- 5.15. Cf. Prob. 5.14 above.
- 5.16. Assume it to be reducible, then consider the homogeneity of the polynomial.
- *5.17. Cf. Th. 5.2.1.12.
- *5.18-21. Cf. Prob. 5.17 and Th. 5.2.1.12.
- *5.22. $k = -5^{5m-1}r^5 - 5^m r$, where m is a positive integer and r is any integer which is not divisible by 5.
- 5.23. $x^3 + (p^3 + 3pq + 3r)x^2 + (3r^2 - 3pqr + q^3)x + r^3 = 0$
- 5.24. Cf. Prob. 27 of §5.2.3.
- 5.25. (i) $p_n y^n + p_{n-1} y^{n-1} + \dots + p_1 y + 1 = 0$
 (ii) $y^n - p_1 y^{n-1} + p_2 y^{n-2} - \dots + (-1)^n p_n = 0$
 (iii) $y^n + k p_1 y^{n-1} + k^2 p_2 y^{n-2} + \dots + k^{n-1} p_{n-1} y + k^n p_n = 0$
 (iv) $(y-m)^n + p_1 (y-m)^{n-1} + \dots + p_{n-1} (y-m) + p = 0$

- 5.26. By hypothesis $f(x) = f(1/x)$, which immediately yields $a_0/a_n = a_1/a_{n-1} = \cdots = a_{n-1}/a_1 = a_n/a_0$, leading to the desired conclusion.
- 5.27. Let q/p , where $p > 0$ and $(p, q) = 1$, be a rational root of the equation; then $(q^2/p^2) + a(q/p) + b = 0$, i.e. $q^2 + apq + bp^2 = 0$, or $-q = p(aq + bp)$, which implies $p \mid q^2$ if $p > 1$, hence $p \mid q$, contradicting the assumption; hence $p = 1$, yielding the desired conclusion.
- 5.28. $3 + \sqrt[3]{3}c - \sqrt[3]{9}$, $3 + \sqrt[3]{3}c - \sqrt[3]{9}c^2$, $3 + \sqrt[3]{3}c^2 - \sqrt[3]{9}c$, where c is the imaginary cubic root of 1.
- 5.29. $-3, 5, -1 \pm i$.
- 5.30. (i) $D = -(4p^3 + 27q^2)$, (ii) $D = (4(p^2 + 12r)^3 - (2p^3 - 72pr + 27q^2)^2)/27$.
- 5.31. $ah - 3bg + 3cf - de = 0$
- 5.32. $y^5 + y^4 - 4y^3 - 3y^2 + 3y + 1 = 0$
- *5.33. Use $x_1 + x_2 + \cdots + x_n = na + (n(n-1)/2)d = -p_1$. Cf. Th. 5.2.3.4.
- 5.34. $x^{n-1} - nx^{n-2} + (n(n-1)/2!)x^{n-3} - \cdots + (-1)^{n-1}n = 0$
- 5.35. $f(x) = (x-r)g(x)$, yielding at once $f(m) = (m-r)g(m)$, where $f(m), m-r, g(m) \in I$.
- 5.36. 3.
- 5.37. If x_1, x_2, x_3, x_4 are the given four roots, then by hypothesis
$$(x_1x_2 - x_3x_4)(x_1x_3 - x_2x_4)(x_1x_4 - x_2x_3) = 0$$
which is symmetric; apply, therefore, Th. 5.2.3.4, to get the desired result.
- 5.38. Cf. Df. 5.2.2.1a, and consider the problem in terms of permutations.
- 5.39. Cf. Th. 4.1.2.5.19.
- 5.40. Prove, first, that $F[a, b]$ is simple when a and b are algebraic over F , i.e. $F[a, b] = F[c]$ for some c algebraic over F ; then generalize.
- *5.41. Since $F[\sqrt{a}]$ has elements of the form $b + c\sqrt{a}$, where $b, c \in F$, let
$$g(x) = x^m + (b_1 + c_1\sqrt{a})x^{m-1} + \cdots + (b_m + c_m\sqrt{a})$$
and consider
$$\tilde{g}(x) = x^m + (b_1 - c_1\sqrt{a})x^{m-1} + \cdots + (b_m - c_m\sqrt{a})$$
- 5.42. Cf. Th. 5.3.1.15.
- 5.43. The first part of the theorem is readily verified, and since R is the minimal subfield of F , any number field containing $\sqrt{3}$ must contain every number of the given form.
- 5.44. Let $p/q = a + b\sqrt{2}$, where $a, b \in R$, and take $u, v \in R$ such that $|a-u| \leq 1/2$, $|b-v| \leq 1/2$; then $s = u + v\sqrt{2}$ is an integer. Since $r = p - qs$ is also an integer and $p/q - s = (a-u) + (b-v)\sqrt{2}$, it follows
$$|Nr| = |Nq \cdot N((p/q) - s)| = |Nq| |(a-u)^2 - 2(b-v)^2| \leq |Nq|/2 < |Nq|$$
- 5.45. If $(c + d\sqrt{m})/2 = (c-d)/2 + d(1 + \sqrt{m})/2$ is a rational integer, so are $(c-d)/2$ and d , since $c \equiv d \pmod{2}$. Conversely, if a and b are any rational integers, then
$$a + b(1 + \sqrt{m})/2 = ((2a+b) + b\sqrt{m})/2$$
which is a rational integer, since $2a+b \equiv b \pmod{2}$.
- 5.46-47. Cf. Df. 5.3.2.13.
- 5.48-51. Cf. Df. 5.3.2.11.
- 5.52. Cf. Th. 5.3.2.12.

INDEX

- Abel's theorem, 284
- Abelian group, 66
- Absolute value, 143
 - of complex numbers, 237
 - in ordered domains, 143
 - of quaternions, 176
- Absorption law, 50, 56
- Abstract group (i.e. group in general), 65
- Addition (*see* Disjunction, Join, Sum, Union), 32
 - of complex numbers, 236
 - of (Dedekind) cuts, 220
 - of matrices, 181
 - of natural numbers, 147
 - of polynomials, 165
 - of quaternions, 175
 - in quotient field, 160
 - of sets, 40
 - of vectors, 179
- Additive group, 66, 69, etc.
- Additive inverse, 69, 131
- Adjoint, 184
- Adjunction, 165, 301
- Aleph-null, 26
- Aleph-one, 26
- Algebraic
 - element, 301
 - extension, 301
 - integer, 313
 - number, 312
- Algebraically complete (closed), 312
- Alternating group, 75
- Anti-automorphism, 176
- Anti-symmetry, 49
- Archimedean (ordered), 152, 161, etc.
- Argument, 237
- Associate, 148
- Associative law, 32
 - for Boolean algebras, 58
 - for fields, 159
 - for groups, 65
 - for lattices, 50
 - for matrices, 188
 - for quaternions, 177
 - for rings, 131
 - for sets, 41
 - for transformations, 37
 - for vectors, 185
- Automorphism, 36
 - anti-, 176
 - identity, 214
 - inner, 106
 - order, 214
 - outer, 106
 - reciprocal, 176
- Axiom
 - of choice, 51
 - of completeness, 222
 - of extension, 24
- Basis
 - for Abelian groups, 124
 - for finite extensions, 302
 - of vector spaces, 181
- Bijjective transformation (*see* Transformation)
- Binary
 - connective, 2
 - operation, 31
- Binomial
 - equation, 282
 - theorem, 27
- Boolean
 - algebra, 56
 - ring, 139
- Buniakovski's inequality (*see* Cauchy-Schwarz inequality)
- Cancellation law,
 - additive, 132
 - for congruences, 149
 - for groups, 66
 - for integral domains, 141
 - multiplicative, 141
- Cantor sequence, 222
- Cap (*see* Meet)
- Cardano's theorem, 287
- Cardinal number, 5
- Cartesian product (*see* Direct product)
- Cauchy-Cantor sequence (*see* Cantor sequence)
- Cauchy-Schwarz inequality (*see* Schwarz inequality)
- Cayley table (cf. Multiplication table)
- Cayley's theorem, 84
- Cell, 42
- Centralizer, 101
- Center of a group, 102
- Chain rule (*see* Syllogism principle)
- Characteristic, 142
- Circuit (designs),
 - parallel, 56, 61
 - series, 56, 62
- Class, 41
 - equation, 96
- Closure, 32, 65, 131, etc.
- Coefficient, 165
- Cofactor, 184
- Column matrix, 181
- Commutative
 - group (*see* Abelian group)
 - ring, 131
 - sfield, 175
- Commutative law, 32
 - for Boolean algebras, 58
 - for fields, 159
 - for groups, 66
 - for lattices, 50
 - for matrices, 188
 - for quaternions, 177
 - for rings, 131
 - for sets, 41
 - for vectors, 185
- Complement
 - in a Boolean algebra, 56
 - in a lattice, 50
 - of a set, 40
- Complete ordered field, 222
- Complex, 66
- Complex number, 236
 - field, 236
 - plane, 237
- Component (or coordinate)
 - of a complex number, 236
 - of a quaternion, 175
 - of a vector, 179
- Composite (number), 148
- Composite (*see* Product)

- Composition
 - index, 122
 - series, 122
- Condition,
 - necessary, 20
 - sufficient, 20
- Conformability, 182
- Congruence, 116, 149f., etc.
- Conjugate
 - complex numbers, 237
 - Gaussian integers, 312
 - quaternions, 176
 - subgroups, 96
- Conjunction, 2
- Connective (*see* Logical connective)
- Constant, 13
 - term, 167
- Contradiction, 3
- Contrapositive (or opposite converse), 10, 21
- Coordinate (*see* Component)
- Correspondence,
 - many-one, 35
 - one-one, 25
- Coset,
 - left, 95
 - right, 95
- Countable (*see* Denumerable)
- Cramer's Rule, 192
- Cubic
 - equation, 283
 - field, 161
- Cup (*see* Join)
- Cut (*see* Dedekind cut)
- Cycles, 74
- Cyclic
 - group, 90
 - permutation, 136

- Decimals, 30
- Decomposition,
 - left, 96
 - right, 96
- Dedekind cut, 219
- Deductive inference (*see* Logical inference)
- Degree
 - of algebraic elements, 303
 - of finite extensions, 302
 - of permutations, 72
 - of polynomials, 181
- De Moivre theorem, 237
- Demonstration, 19
- Denumerable, 26
- Dependence,
 - functional, 272
 - linear, 180, 302
- Determinant, 183
 - of Möbius mappings, 249
- Diagonal matrix (*see* Matrix)
- Difference, 142, 147
 - group, 116
- Dihedral group, 77
- Dimension
 - of indeterminates, 168
 - of vectors, 181
- Direct (or Cartesian)
 - product, 34, 123
 - sum, 203
- Discriminant
 - of cubic equations, 323
 - of quadratic equations, 283
 - of quartic equations, 323
- Disjoint, 42
 - cycles, 75
- Disjunction, 2
- Distributive lattice, 50
- Distributive law, 32
 - for Boolean algebras, 56
 - for fields, 159
- Distributive law (cont.)
 - for integers, 147
 - for lattices, 50
 - for matrices, 188
 - for quaternions, 177
 - for rings, 131
 - for sets, 41
 - for vectors, 179
- Divisibility, 148
- Division algorithm
 - for integers, 148, 202
 - for polynomials over a field, 251
 - for polynomials over a ring, 167
- Division ring, 175
- Divisor, 148
 - zero, 131
- Domain,
 - Gaussian, 252
 - integral, 141
 - of function, 34
- Dot product, 182
- Duality principle, 32

- Eisenstein's theorem, 252
- Elementary symmetric function, 271
- Embedded, 198
- Endomorphism, 36
 - of groups, 105
 - of rings, 132
- Equivalence,
 - of algebras, 56
 - relation, 25
- Euclidean algorithm
 - for Gaussian integers, 324
 - for integers, 149, 156
 - for polynomials, 252
- Euclidean
 - geometry, 21
 - ring, 203
 - (vector) space, 179
- Even permutation, 75
- Existential quantification, 13
- Extension
 - finite, 302
 - multiple algebraic, 302
 - simple algebraic, 302
 - transcendental, 302

- Factor, 148
 - group, 115
 - ring, 206
- Factor theorem, 168
- Fermat's theorem, 117, 211 (Prob. 4.17)
- Ferrari-Euler's theorem, 283
- Field, 159
 - algebraic (number), 311
 - Archimedean ordered, 161
 - complete ordered, 222
 - complex number, 236
 - cubic, 161
 - Gaussian number, 312
 - noncommutative, 175
 - number, 160
 - ordered, 161
 - prime (or minimal), 159
 - quasi-, 175
 - quotient, 160
 - rational number, 214
 - real number, 221
 - skew, 175
 - sub-, 159
- Finite
 - extension, 302
 - field, 160
 - group, 67
 - induction, 33
 - set, 26
- Finite induction principle, 33
- Form, 166

- Four group, 76
- Fraction, 160
- Function, 34
 - polynomial, 166
- Functional
 - calculus, 13
 - dependence, 272
 - independence, 272
- Fundamental Theorem
 - of algebra, 281
 - of arithmetic, 149
- Galois field, 160
- Gauss' theorem, 252
- Gaussian
 - domain, 252
 - integer, 313
 - number, 312
- Generalization principle, 15
- Generator
 - of field extensions, 301
 - of groups, 90
 - of ideals, 202
 - of vector spaces, 180
- Greatest common divisor (g.c.d.), 149
- Greatest lower bound (g.l.b.), 49, 221
- Group, 65
 - Abelian (or commutative), 66
 - additive Abelian, 66
 - alternating, 75
 - composition-quotient, 122
 - demi-, 66
 - difference, 116
 - dihedral, 77
 - factor, 115
 - finite, 67
 - four, 76
 - Hamiltonian, 110
 - infinite, 67
 - loop, 66
 - monoid, 66
 - octic, 108-9
 - of quaternions, 108, 110, 115
 - of transformations, 72
 - of translations, 77
 - permutation, 73
 - quasi-, 66
 - quotient, 115
 - semi, 66
 - simple, 110
 - symmetric (permutation or substitution), 74
 - sub-, 66
- Groupoid, 31
- Hamilton
 - group, 116
 - number couple, 236
 - quadruple, 175
- Height, 29
- Homogeneous polynomial, 168
- Homographic mapping, 238
- Homomorphism, 35
 - improper, 206
 - of groups, 83
 - of rings (and similarly, of domains, fields, etc.), 132
 - proper, 206
- Hypercomplex number, 179
- Ideal, 201
 - improper, 202
 - left, 202
 - maximal, 207
 - prime, 206
 - principal, 202
 - proper, 202
 - right, 202
 - two-sided, 202
- Ideal (cont.)
 - unit, 202
 - zero, 202
- Idempotent law
 - for Boolean algebras, 57
 - for lattices, 50
 - for sets, 41
- Identity, 33
 - element, 50, 56, 65, 131, 159, 187, etc.
 - of indiscernibles principle, 1
 - left, 65
 - matrix, 183
 - operation, 105
 - permutation, 74
 - right, 65
 - transformation, 37
- Image, 34
- Imaginary part (or component), 236
- Inclusion, 24, 56
- Independence
 - functional, 272
 - linear, 180, 302
- Indeterminate, 165
- Index, 96
 - composition, 122
- Induction principle, 33
- Inequality
 - Schwarz, 143
 - triangle, 143
- Inference principle, 4
- Infimum (*see* Greatest lower bound)
- Infinite
 - characteristic, 142
 - group, 67
 - set, 26
- Injective transformation, 34
- Inner automorphism, 106
- Inner product, 182
- Integers, 147
 - algebraic, 313
 - rational, 147
- Integral domain, 141
- Intersection, 40
- Into, 35
- Invariant subgroup, 110
- Inverse, 33
 - element, 50, 56, 65, 131, 159, 187, etc.
 - left, 65
 - matrix, 185
 - permutation, 74
 - right, 65
 - transformation, 37
- Irrational number, 30, 220
- Irreducible polynomial, 252
- Isomorphism, 35
 - of groups, 84
 - of rings (integral domains, fields, etc.), 132
- Join, 40
 - of classes, 31-2
 - of cosets, 96
 - of subsets, 40
- Joint denial, 9
- Jordan-Hölder's theorem, 123
- Kernel, 111
- Klein's group (*see* Four group)
- Kronecker delta, 184
- Lagrange's theorem, 96
- Lattice, 50
 - Boolean, 50
 - complemented, 50
 - distributive, 50
 - modular (or Dedekind), 50
- Leading coefficient, 167
- Least common multiple (l.c.m.), 149
- Least upper bound (l.u.b.), 49, 221

- Left
 - coset, 95
 - decomposition, 96
 - ideal, 202
 - identity, 65
 - inverse, 65
- Linear
 - combination, 149, 180
 - dependence, 180, 302
 - independence, 180, 302
 - space, 179
 - sum, 180
- Lower bounds, 49, 221
- Map, 34
- Mapping (*see* Transformation)
- Mathematical induction, 33
- Matric product, 182
- Matrix, 181
 - column, 181
 - diagonal, 183
 - identity, 183
 - multiplication, 182
 - nonsingular, 184
 - row, 181
 - scalar, 183
 - square, 183
 - sub-, 184
 - zero, 183
- Maximal
 - ideal, 207
 - normal subgroup, 122
- Meaning, 1
- Mean-value theorem, 281
- Meet, 40
 - of classes, 41-2
 - of sets, 40
 - of subfields, 200
- Metatheorem (*see* Introduction)
- Minimal
 - field, 159, 198
 - generating system, 124
- Minor, 184
- Möbius (or linear or homographic) mapping, 238
- Modular lattice, 50
- Module, 66
- Modulus, 149, 237
- Monic polynomial, 167
- Monoid, 66
- Multinomial Theorem, 331
- Multiple, 148
 - algebraic extension, 302
- Multiplication table, 75
- Multiplicity, 280
- Negation, 2
- Negative inference (or contrapositive) principle, 4
- Nilfactor, 189
- Norm
 - of Gaussian integers, 318
 - of Gaussian numbers, 312
 - of quaternions, 176
- Normal subgroup, 110
- Normalizer, 101
- Null
 - operator, 105
 - set, 25
 - space, 181
- Number field,
 - algebraic, 311
 - complex, 236
 - rational, 214
 - real, 221
- Odd Permutation (*cf.* permutation)
- One-to-one (or one-one) correspondence, 25, 35
- Onto, 35
 - into (*i.e.* onto and into), 35, 84, etc.
 - or into, 83, etc.
- Operand, 31
- Operator, 31, 105
- Order
 - automorphism, 214
 - isomorphism, 221
 - of an element of a group, 90
 - of a group, 67
 - of a square matrix, 183
- Ordered
 - Archimedean, 161
 - domain, 142
 - field, 221
 - partly, 49
- Ordering, 50
 - partial, 51
 - simple, 51
 - well-ordering, 33, 51
- Outer automorphism, 106
- Parallel, 21
- Partial fraction, 252
- Partition, 42
- Peano axioms, 146
- Permutation, 73
 - circular, 74
 - cyclic, 73
 - even, 75
 - identity, 74
 - inverse, 74
 - odd, 75
 - symmetric, 74
- Polynomial, 165
 - elementary symmetric, 271
 - function, 166
 - homogeneous, 168
 - irreducible, 252
 - monic, 167
 - prime, 252
 - primitive, 252
 - reducible, 252
 - relatively prime, 252
 - symmetric, 270
- Positive integers, 142
- Prime, 149
 - field, 159
 - ideal, 206
 - relatively, 149
- Primitive (or postulate or axiom), 4
- Principal ideal, 202
- Principia Mathematica, 4
- Product (or composite),
 - direct (or Cartesian), 34, 123
 - dot (or scalar or inner), 182
 - matric, 182
 - of elements of a group, 65
 - of subsets of a set, 95
 - of transformation, 36
 - scalar, 179, 181
- Proof, 19
- Proper
 - divisor, 148
 - homomorphism, 206
 - ideal, 202
 - subgroup, 66
 - subset, 25
- Proposition (or statement), 1
 - existential, 13
 - universal, 13
- Quadratic
 - equation, 283
 - field, 161
- Quantification, 13
- Quantifier,
 - existential, 13
 - universal, 13
- Quartic equation, 283

- Quasi-
 - field (*see* Sfield)
 - group, 66
- Quaternion group (*see* Group)
- Quaternions, 175
- Quintic equation, 284
- Quotient, 160, 167, 251
 - field, 160
 - group, 115
 - ring, 206
- Radical (root extraction), 283
- Range, 34
- Rational
 - form, 257
 - integer, 147
 - number, 29, 214
 - number field, 214
- Real
 - number field, 221
 - numbers, 29, 221
 - part (component), 236
- Reciprocal automorphism, 176
- Reducible polynomial, 252
- Referent, 34, 49
- Reflection, 77
- Reflexive law, 25
- Relation, 34
- Relative prime, 149
- Relatum, 34, 49
- Remainder class, 116
- Remainder theorem, 168, 251
- Residue class, 116
 - domain, 142
 - group, 116
 - ring, 135, 198
- Right
 - coset, 95
 - decomposition, 96
 - ideal, 202
 - identity, 65
 - inverse, 65
- Ring, 131
 - adjunction, 165
 - commutative, 131, 139
 - division, 175
 - noncommutative, 131, 175
 - principal ideal, 202
 - sub-, 132
 - with unity, 131
 - with zero-divisors, 131
- Root (or zero), 168, 251
 - of cubic equations, 283
 - of quadratic equations, 283
 - of quartic equations, 283
- Rotation, 76
- Row matrix, 181
- Scalar
 - matrix, 183
 - multiplication, 179, 181
 - product, 179, 181
- Schwarz inequality, 143
- Sequence,
 - Cauchy-Cantor, 222
 - logical, 19
- Set, 24
- Sfield, 175
- Sign, 1
 - Descartes' rule, 282
- Simple extension, 302
- Singleton, 24
- Skew field, 175
- Spanned (or generated), 180
- Specialization principle, 15
- Square matrix, 183
- Subdomain, 141
- Subfield, 159
 - minimal (or prime), 159, 198
- Subgroup, 66
 - common, 90
 - conjugate, 110
 - cyclic, 90
 - invariant, 110
 - maximal normal, 122
 - normal, 110
 - self-conjugate, 110
- Subring, 132, 198
- Subspace, 180
- Substitution principle, 3
- Successor, 147
- Sum (*see* Join, etc.)
 - direct, 203
 - linear, 180
- Supremum (*see* Least upper bound)
- Surjective transformation, 35
- Syllogism principle, 5
- Symmetric, 25
 - anti-, 49
 - group, 74
 - polynomial, 270
- Tautology, 3
- Total matrix algebra, 185
- Transcendental, 301
- Transform, 31, 96
- Transformation, 34
 - bijective, 35
 - injective, 35
 - linear, 238
 - surjective, 35
- Transforming element, 96
- Transitive, 25
- Translation, 77
- Transpose, 184
- Transposition, 75
- Triangle inequality, 143
- Trichotomy, 142
- Truth table, 2
- Unary operation, 31
- Union (*see* Join, etc.)
- Unique factorization theorem, 149, 252
- Unit, 148
- Unit set, 24
- Unity, 131
- Universal
 - bounds, 50
 - proposition, 13
 - quantifier, 13
- Upper bounds, 49, 221
- Variable, 13
 - bound, 13
 - free (flagged), 13
- Vector, 179
 - space, 179
- Venn diagram, 14
- Vierergruppe (*see* Four group)
- Well-ordering principle, 33
- Wronskian, 194
- Zermelo's postulate, 33
- Zero
 - divisor, 131
 - ideal, 202
 - matrix, 183
 - subspace, 180
 - vector, 180
- Zorn's lemma, 51

